DOCUMENT RESUME

ED 397 076 TM 025 071

AUTHOR Kaplan, Randy M.; And Others
TITLE Evaluating a Prototype Essay Scoring Procedure Using Off-the-Shelf Software.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-95-21
PUB DATE Jul 95
NOTE 81p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Computer Software; *Constructed Response; *Essay Tests; Grammar; *Scoring; *Theory Practice Relationship
IDENTIFIERS Commercially Prepared Materials; *Decision Models; *Grammar Checkers; Test of English as a Foreign Language

ABSTRACT

The increased use of constructed-response items, like essays, creates a need for tools to score these responses automatically in part or as a whole. This study explores one approach to analyzing essay-length natural language constructed-responses. A decision model for scoring essays was developed and evaluated. The decision model uses off-the-shelf software for grammar and style checking of the English language. The best performing grammar checking programs from among several commercial programs were selected to construct a decision model for scoring the essays. Data produced from the selected grammar programs were used to make a decision about the score for an essay. Through statistical and linguistic methods, the performance of the decision model was analyzed in an effort to understand its usefulness and practicality in a production scoring setting. A sample of 80 essays was selected from Test of Written English essays prepared for the Test of English as a Foreign Language. Using four grammar-checking programs, 320 analyses were produced. Results indicated that a model could be constructed using the commercial programs and that about 30% of the essays could be scored correctly. Scores derived from the scoring model could be accepted as accurate, but the number of essays scored does not yet warrant its application in a practical setting. Three appendixes contain sample grammar check outputs, a categorization of errors from the grammar checkers, and essay analysis data. (Contains 16 tables, 5 figures, and 6 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-95-21

ED 397 076

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

EVALUATING A PROTOTYPE ESSAY SCORING PROCEDURE USING OFF-THE-SHELF SOFTWARE

Randy M. Kaplan
Jill Burstein
Harriet Trenholm
Chi Lu
Donald Rock
Bruce Kaplan
Susanne Wolff

BEST COPY AVAILABLE

Educational Testing Service
Princeton, New Jersey
July 1995

TM025071

Evaluating a Prototype Essay Scoring Procedure Using Off-The-Shelf Software

Randy M. Kaplan, Jill Burstein, Harriet Trenholm, Chi Lu, Donald Rock, Bruce Kaplan, and Susanne Wolff

April 27, 1995

This work was carried out under the auspices and support of the Program Research Planning Council (PRPC), Project No.: 968-21.

Copyright © 1995. Educational Testing Service. All rights reserved.

Abstract

Constructed-response items, whose responses consist of words, phrases, sentences, paragraphs, and essays are among the most difficult and costly to score. The increased use of constructed-response items like essays creates a need for tools to partially or fully automatically score these responses. This study explores one approach to analyzing essay-length natural language constructed-responses.

In this study we develop and evaluate a decision model for scoring essays. The decision model uses off-the-shelf software for grammar and style checking of the English language. The first part of this study consisted of an evaluation of several commercial grammar checking programs. From this evaluation we select the best performing grammar checking programs to construct a decision model for scoring the essays. The second part of the study uses data produced from the selected grammar checking program(s) to make a decision about the score for an essay. Through statistical and linguistic methods, we analyze the performance of the decision model in an effort to understand its usefulness and practicality in a production scoring setting.

Evaluating a Prototype Essay Scoring Procedure Using Off-The-Shelf Software

One of the challenges we face in the ongoing evolution of tests from traditional multiple-choice items to the more complex constructed-response items is how to score responses for the latter. As the nature of an item becomes more complex, so does the nature of its response. The increase in complexity translates into increased costs for examinees, related to the increased cost of scoring an examination composed of these complex item types. Since examinations include more complex item types, we must explore new approaches to scoring which include semi- and fully automatic and semi-automatic means for scoring.

An important class of complex item types for which we must explore new scoring methodologies are those whose constructed responses are phrases, sentences, paragraphs, and essays in English or some other natural language. By natural language we mean a language that is used by humans for communication. Scoring natural language responses by traditional methods is a time consuming and costly process. The volume of responses to read and score is formidable enough in scoring short-answer responses. For essays, although the number may be comparatively small, and the relative length of essays to be read from an administration might be small, the number of essays to be read from an administration might prohibit their use in large testing programs. The purpose of this study is to explore how we might reduce the work and cost involved in scoring particular types of essays.

An item type used in the Test of Written English (TWE), administered as part of the Test of English as a Foreign Language (TOEFL), requires an examinee to write an essay. The essay is scored holistically on characteristics including grammar, style, and the ability to organize and support ideas. TWE essays are scored on a six point scale. If an essay is rated as a 1 or 2 on this scale, we can infer that the examinee's competence in using grammar, formulating style, and organizing written material is low. If, on the other hand, an examinee's essay is given a rating of 5 or 6, we can assume that the skills in these abilities are very good. Our research originally focused on developing a procedure for classifying essays into two groups: those essays whose score would be a 1 or 2 and all other essays. Later, we expanded the classification so that essays would be classified into three groups: those which are rated a 1 or 2, those which are rated a 5 or 6, and all other essays (those which are rater 3 and 4).

Significant expense can be incurred in any project that requires the creation of a complex software program. Rather than create such a program for this project, and incur the related expense, part of this study is to evaluate the possibility of using commercially available software for processing essays and ultimately producing essay scores. For this project, we used four commercially available grammar and style checking programs to analyze essays.

Our goal for this project was to create a model of categorizing essays into groups based on the features of the essays as produced by the grammar-

checking programs. Our hypothesis can be stated as follows: An essay receiving a particular score on the six point scale will have a set of identifiable characteristics that can be recognized by a grammar-checking program associated with it. To develop a scoring model, and test this hypothesis, we analyzed a sample of essays (n=300), and collected analyses from the grammar and style checkers. We then normalized these analyses so that the results of one grammar-checking program could be related to the results of another.

Background

Very little research has been published which discusses potential capabilities and applications for computer-based essay scoring. This section briefly reviews the most recently published work in this area. This short review is intended to provide the reader with some background and perspective about this virtually unexplored area.

The most recently published work with regard to computer-based scoring of essays was Page and Petersen (1995). This article is an update of Page's Project Essay Grading (PEG) system originally talked about in Page (1966). Page and Petersen claim that correlations between PEG and human graders were higher than correlations between human graders. In the Page and Petersen study, 1,314 PRAXIS essay items were provided by ETS so that they could be scored by the PEG system. All of these essays had been scored by 2 human graders. The essays were randomly divided into a test set of 300

essays and a research set of 1,014. They claim that the research set was used "...formatively to fine-tune the computer program..." However, the article barely touches on what procedures are used in general to score essays. The authors do mention a variable they use called a *prox* (approximations). Unfortunately, the only example which they provide of a prox is essay length. Certainly, essay length alone is too crude a measure to accurately predict essay scores. What is actually done in the fine-tuning process is never revealed. Since the authors claimed that correlations between human judges are generally no higher than .50 or .60, ETS provided 4 extra human grader scores for a random 300 of the 1,014 essays in the research set, and for the 300 test essays, so that there were a total of 6 human grader scores for 600 essays. Page and Petersen claim that for the 300 test essays , the mean correlation between the computer and the 6 human judges was .742, as compared to the mean correlation between the six judges which was .646; the mean correlation between the computer and pairs of human judges was .816, while the mean correlation between the pairs of human judges was .761; and, the mean correlation between the computer and three human judges was .846, and the mean between the judges was .834. The article never states what variable the correlations are based on.

Though the reported results of this work appear to be promising, at least on the surface, the article does not document how any of the results were derived. That is, the article never explains the machine-based

procedures which were implemented in order for PEG to successfully score essays. This work requires more discussion about PEG's scoring procedures before the reliability of this system can be fairly assessed.

The Test of Written English

The Test of Written English (TWE) is a constructed response item that is part of the Test of English as a Foreign Language (TOEFL). Examinees are given thirty minutes to compose, write, and revise an essay about a particular topic. They are told that their essays will be judged on overall quality. An example of a TWE essay item is shown in Figure 1 (TOEFL, 1989).

Figure 1 - Sample TWE Essay Item
Supporters of technology say that it solves problems and makes life better. Opponents argue that technology creates new problems that may threaten or damage the quality of life. Using one or two examples, discuss these two positions. Which view of technology do you support? Why?

Two essay responses are shown in Figures 2 and 3. The first of these was assigned a score of 1 and the second a score of 6.

Figure 2 - Sample TWE essay response scored 1 on a scale of 6

Now a days in the life of the technology it solves problems. But damage the quality of the life if very important. Because the many people to the quality of life is very high than the yesterday socizat. They are use it buys goods is more good than yestersay. To the many people to need the high quality are too many.

Figure 3 - Sample TWE essay response scored 6 on a scale of 6

There are several viewpoints on the implications of technological change and advancement and such schools of thought which considerably vary have their respective validity. Technological change has its advantage and disadvantages. For one, it is true that it partly solves problems and makes life better. At the same time, technological chnages may likely create new problems thereby threatening or damaging quality of life.

In the developing economics, for instance, technological advantages has both its merits and demerits. The introduction and seeming acceptability and usefulness of computers have somehow helped increase the efficiency of several firms. It is not only in the insdustrial sector that technological change proven to be very effective. In the agricultural sector, for example, the introduction of new technologies in increasing production has been very effective in expanding agricultural produce. These are just a few examples to *illustrate the advantages of technological advancement.

On the other hand, countries should be more careful on their choice of technology since it must be noted that while certain types of technology are adaptable to developed economies the same type of technology may not fit the envisionment of developing conuntries due to differeing economic, social, cultural, and political factors. For example, infrastructure improvements such as construction of irrigation dam in the mountains of the Phillipines where several natives reside may likely be resisted by the population due to cultural factors. They may prefer not to have such improvements in view of traditional values. Another example is the pollution impact of some technological improvements particularly in the industrial sectors.

The choice and adaptability of new tecgnology should therefore be carefully studied. The short, medium, and long term impact of such technology is very important particularly for developing economies. The benefits should always be greater than the costs.

I am inclined to support both positions because both views have their own validity. However, I am more concerned that technological advancement is really beneficial to countries so long as they are aware of the disadvantages of such technology.

As you can see in Figures 2 and 3, these essays differ markedly in construction, style, and length, etc. If we can categorize the difference

between essays based on their characteristics, we would have a procedure to score essays.

In the TWE program, scoring of a TWE essay is based on a rubric consisting of six categories. As we mentioned, the scale ranges from 1 to 6 and each of the ratings has associated with it specific characteristics that graders are looking for when scoring an essay. The next figure shows the criteria for essays assigned a score of 1 and those assigned a score of 6.

Table 1 - TWE essay scoring criteria for scores of 1 and 6

Score 1	Score 6
incoherent undeveloped contains severe and persistent writing errors	effectively addresses the writing task is well organized and well developed uses clearly appropriate details to support a thesis or illustrate ideas displays consistent facility in the use of language demonstrates syntactic variety and appropriate word choice

Software for Grammar and Style Checking of the English Language

Computer-based grammar and style checkers have been available for several years. Two of the oldest commercial products are RightWriter and Grammatik. A third product, named CorrectGrammar, is somewhat newer than both Grammatik and RightWriter. The newest product is one called PowerEdit.[1]

Grammar-checking programs analyze text, and give feedback about writing. The feedback consists of messages that indicate errors in syntax,

[1]Although this is the newest and most sophisticated of the grammar checking programs, it was a short-lived product and is no longer commercially available. Nevertheless, as the most sophisticated, it remains one of the important elements of our analysis.

word usage, and sometimes elements of style. All grammar-checking programs give these kinds of feedback in varying degrees of accuracy and appropriateness. Appendix A contains samples of the analysis produced by each of the grammar-checking programs. The differences between the grammar-checking programs makes comparing the output of one program to another a difficult task.

At the beginning of the study, all four grammar-checking programs were used. Our intention was to find the program that produced the best results in being able to score TWE essays. Although it was our initial belief that the more sophisticated the grammar-checking program is the better able it would be to provide the basis for an accurate essay score, this was by no means something that we knew for sure. Rather than make assumptions about which grammar-checking program would perform best, all four were evaluated.

The complexity of a grammar-checking program can be judged by considering how it analyzes language. Of these four grammar-checking programs, three recognize linguistic patterns (so-called pattern-based analyzers), and the fourth analyzes sentence structure.

Grammatik, RightWriter, and CorrectGrammar are pattern-based grammar-checking programs. These programs consist of large libraries of patterns that represent various kinds of English language sentence constructions. The performance and accuracy of a grammar-checking program based on patterns depends on the number of patterns built into the

program and the ability of the program to match sentences and parts of sentences against the library patterns.

For example, a pattern in a grammar-checking program might be used to determine if a sentence is written in the passive voice. A common problem with a pattern-based approach to grammar-checking is that all too often the patterns apply to a large class of sentences or phrases. This results in an analysis that contains many messages that are incorrect or irrelevant. It is up to the user of the analysis to judge whether a message is relevant or not.

Unlike the other grammar-checking programs, Poweredit bases its analysis on structures produced by parsing sentences. Parsing is a process by which a computer program analyzes a sentence and creates a syntactic structure for the sentence. The result of the parsing process is a parse structure. Basing a grammatical analysis on parse structure may result in a more accurate analysis because the structure produced by the parser are based on the grammar of the language. Whether this is actually true, that a parser-based analysis will yield better analysis results, and therefore better feedback, is a question we investigated in the current study.

Method

A sample of 80 essays was selected at random from a database of TWE essays prepared for TOEFL (Frase, 1991). Each grammar-checking program was used to process an essay. The results of these analyses were collected. A total of 320 analyses were produced. As we mentioned, each of the four

grammar-checking programs produces output and messages that are specific to the program. In order to compare one grammar-checking program with another, it was necessary to find some basis for comparison. We normalized the set of messages produced by all of the grammar-checking programs. Each grammar-checking program can produce a finite set of messages. By collecting these messages and placing similar messages into similar categories, we have a way to compare these grammar-checking programs. A set of categories based on the error classifications produced by the PowerEdit grammar-checking program was used to classify errors from all four grammar-checking programs. The categories used to classify each of the errors are listed and defined in Table 2.

Table 2 - Grammar checker message categories	
Category	**Category description**
balance	this type of message is produced when the length of the subject of the sentence is much greater than the length of the predicate of the sentence.
cohesion	cohesion messages are issued when there is a question about a particular phrase used to connect two sentences.
concision	messages of concision alert a writer to redundancy in a sentence.
discourse	discourse-type messages focus on characteristics of a passage like strength, focus, topic, and clarity.
elegance	elegance messages typically appears when an analyzer makes a recommendation about a particular phrase. For example, an elegance message will be given if a writer uses a vulgar expression.
emphasis	this type of message usually is given when a sentence is written in the passive voice, when a more effective version could have been formulated in the active voice.
grammar	grammar message appear when their are specific identifiable errors in grammar usage. For example, a missing word may result in a grammar message.
logic	messages dealing with logic and flow are classified as logic messages.
precision	a grammar checker will issue a message about precision when it determines that a sentence may be too wordy or that the sentence may have too many possible topics.
punctuation	punctuation messages are produced if a sentence contains a misused punctuation mark.
relation	a "relation" message may be issued when a sentence contains a potential problem in anaphoric reference, or when particular words or phrases are being used in a questionable way in the sentence.
surface	surface messages occur when a sentence contains misspellings, words that are not part of the English language, and sentences that may be confusing to read.
transition	if, in a sentence, an introductory phrase is incorrectly used, or if a clause in the sentence might be placed elsewhere for better readability, a transition message will be produced.

Table 2 (continued) - Grammar checker message categories	
Category	**Category description**
unity	unity messages will occur whenever a word, group of words are used incorrectly, effecting the flow or clarity of the sentence. For example, when a phrase possibly refers to an incorrect phrase, a unity message will be produced.
usage	this type of message will be produced whenever a word or phrase is used incorrectly effecting the grammar of the sentence. For example, a usage message will be produced in the case of a double negative.

Appendix B contains the categorizations of error messages from the grammar checkers. An excerpt from this table is shown in Figure 4.

Figure 4 - Excerpt from grammar checking program error classifications

Category	Error Number in Power Edit	Error Description in Poweredit	Error Message in Poweredit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Cohesion	065	Style/ Writing Style/ Redundant Subjects		29. These words may be redundant; consider omitting them. 30. Redundant expression. Use ... instead.	26. Redundant phrase	S14. Consider omitting: ... U13. Redundant. ... U13. Redundant. Replace ... by ...

As shown in Figure 4, an attempt was made to compare an error message from a grammar-checking program with others that are similar. This process was carried out manually for all error messages produced for all of the essay analyses.[2]

[2] The categroizations of each error message from each grammar checker were made by staff working on the data analysis process. As such, these categorizing of error messages into meta-categories may not be optimal. We did not explore how alternate categorizations affect performance of the scoring process, although, as is presented later in this report, linguistic analysis indicates that it may be inappropriate to use meta-categories.

After the error messages were classified, the number of errors of each error category were calculated. This resulted in a vector of 15 error category counts for each essay. As each grammar-checking program produced one or more errors in each category, an essay analysis record consisted of sixty individual fields[3]: fifteen per grammar-checking program for each of four programs. Appendix C contains the description of the resulting data record used in the model building process.

Regressions were run to see how well a vector of error message scores from a particular grammar-checking program predicted the mean score of an essay calculated from two human raters. This produced the correlations shown in Table 3. The statistics included in this analysis were means, standard deviations, and correlations. The purpose was to identify component scores from each of the four grammar checkers which relate to the TWE mean score for an essay.

Table 3 - Analysis results for first 80 essays

Grammar Checker	multi-correlation	amount of variation explained	probability	number of meta-categories for the grammar checker[4]
Poweredit	.799	.638	.000	15
Grammatik	.582	.339	.001	11
CorrectGrammar	.521	.271	.005	10
RightWriter	.703	.494	.000	10

[3] It is quite possible that a grammar checker could have issued several error messages for the same sentence. This would indicate a possible need to weight the results from a grammar checker in terms of the number of errors produced for any given sentence. This consideration was not included in the present analysis.

[4] In some cases, not all meta-categories were filled by a grammar checker. This column reflects the number of meta-categories used in the regression model.

The correlations[5] between mean score of the human raters and the estimation models were strong enough to continue the analysis by increasing the sample size.

Two samples were used to analyze the model scoring performance. Sample 1 consisted of 461 cases while sample 2 had 475 cases. Mean ratings of the experts were recorded for each essay and used as the outcome variable in the following analysis. Two analytical procedures were used. The ordinary least squares regression (OLS) was used as preliminary screening procedure to identify the better methods for predicting the expert decisions. That is, separate stepwise regression models were used to find the "best" weighted combination of subscores from each of the competing grammar-checking programs for predicting: 1)whether a paper should be classified into one of two categories: either a 1 or 2 paper or a 3 or better paper and 2) whether a paper should be classified as a 5 or better or less than a 5 paper. Thus, the first stage of the next part of the analysis attempted to predict two different dichotomous decisions, one at the lower end of the scoring scale and the other at the upper end of the scale.

The results of this analysis were then taken to a second and final stage where the final prediction models were developed. For the final

[5] As H. Breland indicated to us in a review of this work, holistic scorings of essays have a reliability near .50. In this work we take the reliability of a score produced by one or more human raters as a basis upon which to compare the automated scoring procedure. We did not seek to improve the reliability of ratings given to these essays by human raters.

comparison of the competing models, the logistic regression was used rather than the OLS since OLS regressions do not provide accurate standard errors when a dichotomous dependent variable is used. While the OLS procedures give unbiased estimates of the parameters and are simple and inexpensive to run, they are less appropriate for getting the final results and were thus used only as a screening device in the first stage. In the second and final stage a double cross validation design was used. That is, the logistic regression model was applied to the two most promising grammar-checking programs from stage 1 in the following sequence. Using sample 1 the logistic regression formed the basis for the prediction models with the two best software candidates from the first stage. The parameter estimates from sample 1 were then applied to sample 2 to get an independent estimate of the goodness of fit of the sample 1 model when applied to an independent sample. The same two best grammar-checking program models from stage 1 were also estimated in sample 2, and these parameter estimates were then "crossed" over to sample 1. This addresses the generalizability and the relative stability of the two best competing models across independent samples.

Criteria for selection of the two best models from among the four competing software models in stage 1 included: 1) prediction accuracy as measured by the multiple correlation in both samples and for both dichotomous criteria, and 2) the stability across samples with respect to the

pattern of significant predictor subscales that were chosen by the stepwise procedure.

Final criteria, i.e., the criteria used to compare the two "best" models that survived the stage 1 screening were: 1) agreement between the classification by the grammar-checking programs and the human expert judgment, and 2) traditional statistical significance tests and various statistical indices of the relationship between the dichotomous outcomes and the predicted probability from the software that a paper belongs in one group or the other. The data sets used in the analysis are summarized in Table 4.

Table 4 - model evaluations			
Model(data used to create model)	**N**	**Data** (data used evaluate model)	**N**
sample 1	461	sample 1	461
sample 2	475	sample 2	475
sample 1	461	sample 2	475
sample 2	475	sample 1	461
sample 1+2	936	sample 1+2	936

Results

Table 5 presents the number of essays that fell into the various categories within each sample and for the total group of papers based on the mean rating by the experts. For example 88 papers in sample 1 had a mean score of 2 or less while 373 (283 + 90) had mean scores greater than 2. This dichotomous classification of being in the low-scoring group versus being in the high-scoring group will be referred to as the low-level classification decision (lld). The remaining dichotomous decision is concerned with whether the paper is a high-level paper or not, i.e., has a mean rating of 5 or greater and will be referred to as the high level decision (hld). The question here, of course, is how well can the software scoring procedures reproduce the lld decision and hld decisions of the experts.

Table 5 - Average Scores for Each Sample, and Combined Samples			
Mean Score	**Original Sample**	**Evaluation Sample**	**Combined**
1-2	88 (19%)	43 (10%)	131 (14%)
3-4	283 (61%)	300 (63%)	583 (62%)
5-6	90 (20%)	132 (27%)	222 (24%)
Total	461	475	936

Inspection of Table 5 indicates that 19% of the sample one papers were rated as 2 or below while only 9% of the sample two papers were judged by the raters to be at this level. To a certain extent the prediction of rare events such as the papers falling at or below 2 is a somewhat difficult task for an automated procedure. That is, it is hard to improve on a simple decision rule that simply assigns everybody to the greater than 2 group. Such a simple decision rule would lead to an overall correct classification rate of 81%. However, it would have a 100% misclassification rate for the papers that were actually rated 2 or less. The lld decision is even more rare in sample 2. With respect to the hld decisions in Table 5, the rarity of a paper falling in the 5 or above category is somewhat less in sample 2 than in sample 1.

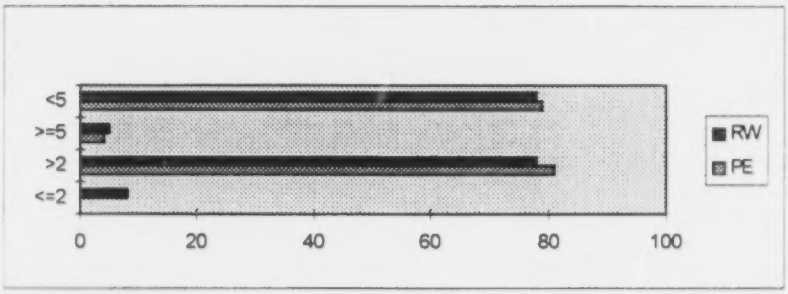
The OLS regression results from the screening stage showed that two of the grammar-checking programs were superior to the other two. The PowerEdit (PE) and RightWriter (RW) grammar-checking programs showed significantly higher multiple correlations and tended to have consistent patterns of statistically significant regression weights associated with the same subscales across both samples. The remaining discussion will center on the comparison of the predictive accuracy of these two procedures for making lld and hld decisions based on the logistic regression.

Table 6 presents the agreement between the expert ratings and that of the logistic regression predicted lld decisions (top half) and hld decisions (lower half) for the PE and RW methods within sample 1. Table 7 presents the parallel results for sample 2. Inspection of Table 6 indicates that while the PE procedure achieved an overall predicted percent correct of 81% by assigning every paper to the greater than 2 group, it misclassified all of the 88 papers that the expert raters classified as being 2 or less. RW, while having a slightly less overall "hit" rate, did much better at the hard task, i.e., making correct assignments of the 2 and less papers. The RW procedure assigned 42% of the "true" 2 or less papers to that category. Clearly RW did a better job of simulating the lld decisions in sample 1 than did PE.

Table 6 - Scoring performance for first sample as model and first sample as data(model: sample 1; data: sample 1)

PE				RW			
Grammar Checker Score				Grammar Checker Score			
LLD	predicted score <= 2	predicted score > 2	total		predicted score <= 2	predicted score > 2	total
mean score <= 2	0 (0%)	88 (19%)	88 (19%)	mean score <= 2	37 (8%)	51 (11%)	88 (19%)
mean score > 2	0 (0%)	373 (81%)	373 (81%)	mean score > 2	13 (3%)	360 (78%)	(373) (81%)
total	0 (0%)	461 (100%)	(81%) (100%)	total	50 (11%)	411 (89%)	461 (100%)
% correctly predicted			81%	% correctly predicted			86%
% of score <= 2 correctly predicted			0%	% of score <= 2 correctly predicted			42%
% of score > 2 correctly predicted			100%	% of score > 2 correctly predicted			97%
Grammar Checker Score				Grammar Checker Score			
HLD	predicted score >= 5	predicted score < 5	total		predicted score >= 5	predicted score <5	total
mean score >= 5	18 (4%)	72 (16%)	90 (20%)	mean score >= 5	24 (5%)	66 (14%)	90 (20%)
mean score < 5	7 (2%)	364 (79%)	371 (80%)	mean score < 5	13 (3%)	358 (78%)	371 (80%)
total	25 (5%)	436 (95%)	461 (100%)	total	37 (8%)	424 (92%)	461 (100%)
% correctly predicted			83%	% correctly predicted			83%
% of score >= 5 correctly predicted			20%	% of score >= 5 correctly predicted			27%
% of score < 5 correctly predicted			98%	% of score < 5 correctly predicted			96%

Table 6a - Summary of scoring performance showing accurate scoring for LLD and HLD decisions for first sample as model and first sample as data(model: sample 1; data: sample 1)



Inspection of the lower half of Table 6 shows that both methods achieved the same overall agreement (83%) between expert and predicted classification for the hld decision, but RW showed a slightly better percentage (27% vs. 20%) in classifying the "true" 5 and over papers.

Table 7 - Scoring performance for second sample as model and second sample as data (model: sample 2; sample: sample 2)

PE				RW			
Grammar Checker Score				Grammar Checker Score			
LLD	predicted score <= 2	predicted score > 2	total		predicted score <= 2	predicted score > 2	total
mean score <= 2	0 (0%)	43 (9%)	43 (9%)	mean score <= 2	21 (4%)	22 (5%)	43 (9%)
mean score > 2	0 (0%)	432 (91%)	432 (91%)	mean score > 2	9 (2%)	423 (89%)	432 (91%)
total	0 (0%)	475 (100%)	475 (100%)	total	30 (6%)	445 (94%)	475 (100%)
% correctly predicted			91%	% correctly predicted			93%
% of score <= 2 correctly predicted			0%	% of score <= 2 correctly predicted			49%
% of score > 2 correctly predicted			100%	% of score > 2 correctly predicted			98%
Grammar Checker/Score: PE >=5				Grammar Checker/Score: RW >= 5			
HLD	predicted score >= 5	predicted score < 5	total		predicted score >= 5	predicted score < 5	total
mean score >= 5	46 (10%)	86 (18%)	132 (28%)	mean score >= 5	35 (7%)	97 (20%)	132 (28%)
mean score < 5	23 (5%)	320 (67%)	343 (72%)	mean score < 5	26 (5%)	317 (67%)	343 (72%)
total	69 (15%)	406 (85%)	475 (100%)	total	61 (13%)	414 (87%)	475 (100%)
% correctly predicted			77%	% correctly predicted			74%
% of score >= 5 correctly predicted			35%	% of score >= 5 correctly predicted			27%
% of score < 5 correctly predicted			93%	% of score < 5 correctly predicted			92%

Table 7a - Summary of scoring performance showing accurate scoring for LLD and HLD decisions for second sample as model and second sample as data (model: sample 2; sample: sample 2)

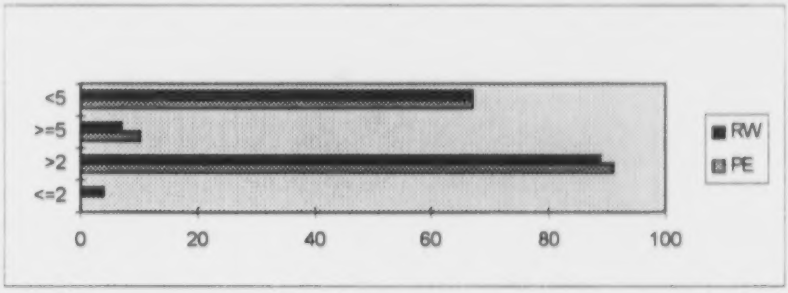


Table 7 presents the parallel analysis carried out on sample 2. The top half of Table 7 indicates that for the lld RW did much better than PE by correctly classifying 49% of the 2 or less papers compared to 0% for PE. For the hld decision (bottom half of Table 7) PE correctly classified slightly more papers in the "true" 5 or greater category than did RW.

Table 8 - Scoring performance for first sample as model and second sample as data (model: sample 1; data: sample 2)

PE				RW			
Grammar Checker Score				Grammar Checker Score			
LLD	predicted score <= 2	predicted score > 2	total		predicted score <= 2	predicted score > 2	total
mean score <= 2	0 (0%)	43 (9%)	43 (9%)	mean score <= 2	7 (1%)	36 (8%)	43 (9%)
mean score > 2	0 (0%)	432 (91%)	432 (91%)	mean score > 2	32 (7%)	400 (84%)	432 (91%)
total	0 (0%)	475 (100%)	475 (100%)	total	39 (8%)	436 (92%)	475 (100%)
% correctly predicted			91%	% correctly predicted			86%
% of score <= 2 correctly predicted			0%	% of score <= 2 correctly predicted			16%
% of score > 2 correctly predicted			100%	% of score > 2 correctly predicted			93%
Grammar Checker/Score: PE >=5				Grammar Checker/Score: RW >= 5			
HLD	predicted score >= 5	predicted score < 5	total		predicted score >= 5	predicted score < 5	total
mean score >= 5	32 (7%)	100 (21%)	132 (28%)	mean score >= 5	40 (8%)	92 (19%)	132 (28%)
mean score < 5	27 (6%)	316 (67%)	343 (72%)	mean score < 5	43 (9%)	300 (63%)	343 (72%)
total	59 (12%)	416 (88%)	475 (100%)	total	83 (17%)	392 (83%)	475 (100%)
% correctly predicted			73%	% correctly predicted			72%
% of score >= 5 correctly predicted			24%	% of score >= 5 correctly predicted			30%
% of score < 5 correctly predicted			92%	% of score < 5 correctly predicted			87%

29

Table 8a - Summary of scoring performance showing accurate scoring for LLD and HLD decisions for first sample as model and second sample as data (model: sample 1; data: sample 2)

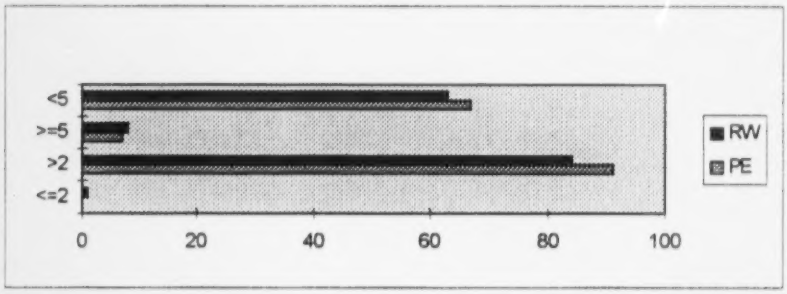


Table 8 presents cross-validation results. The equation developed on sample 1 is applied to sample 2 data. As pointed out above, this is a much more rigorous test of the stability of the prediction models across independent samples. Inspection of the top half of Table 8 (lld) and the bottom half of Table 8 (hld) indicates that RW did somewhat better in classifying papers into both the low level classification and the high level classification.

It should be pointed out that while RW seems superior to PE, the two checkers make different sorts of misclassifications. If, for example, classifying a high-scoring essay as a 1 or 2 is a more serious error than classifying a low-scoring essay as a 3 or greater, then one might prefer PE for lld decisions.

Table 9 - Scoring performance for second sample as model and first sample as data (model: sample 2; data: sample 1)

PE				RW			
Grammar Checker Score				Grammar Checker Score			
LLD	predicted score <= 2	predicted score > 2	total		predicted score <= 2	predicted score > 2	total
mean score <= 2	0 (0%)	88 (19%)	88 (19%)	mean score <= 2	26 (6%)	62 (13%)	88 (19%)
mean score > 2	0 (0%)	373 (81%)	373 (81%)	mean score > 2	6 (1%)	367 (80%)	373 (81%)
total	0 (0%)	461 (100%)	461 (100%)	total	32 (7%)	429 (93%)	461 (100%)
% correctly predicted			81%	% correctly predicted			85%
% of score <= 2 correctly predicted			0%	% of score <= 2 correctly predicted			30%
% of score > 2 correctly predicted			100%	% of score > 2 correctly predicted			98%
Grammar Checker/Score: PE >=5				Grammar Checker/Score: RW >= 5			
HLD	predicted score >= 5	predicted score < 5	total		predicted score >= 5	predicted score < 5	total
mean score >= 5	28 (6%)	62 (13%)	90 (20%)	mean score >= 5	22 (5%)	68 (14%)	90 (20%)
mean score < 5	20 (4%)	351 (76%)	371 (80%)	mean score < 5	12 (3%)	359 (78%)	371 (80%)
total	48 (10%)	413 (90%)	461 (100%)	total	34 (7%)	427 (93%)	461 (100%)
% correctly predicted			82%	% correctly predicted			83%
% of score >= 5 correctly predicted			31%	% of score >= 5 correctly predicted			24%
% of score < 5 correctly predicted			95%	% of score < 5 correctly predicted			97%

Table 9a - Summary of scoring performance showing accurate scoring for LLD and HLD decisions for second sample as model and first sample as data (model: sample 2; data: sample 1)

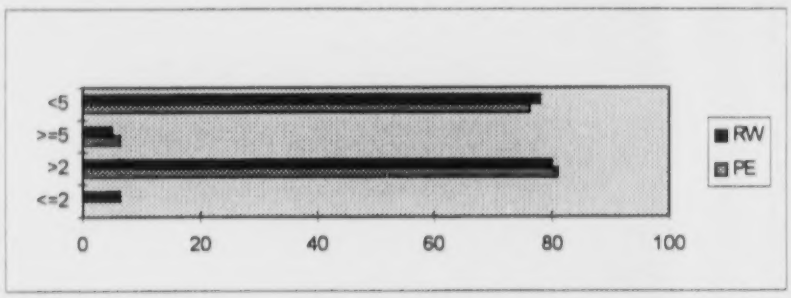


Table 9 presents the results for prediction models developed in sample 2 and cross-validated to sample 1. The results for lld are quite similar to the those found in the other cross-validation. That is, RW is better at classifying the llds, than is PE, subject to the utilities one wishes to assign to the different errors. For the hlds PE appears to do a slightly better job. On the whole, however, RW not only appears to do as good a job or better than PE, but also appears to be at least as stable, if not more stable, as indicated by the cross-validations.

Table 10 - Scoring performance for combined sample as model and combined sample as data (model: combined; sample: combined)

PE				RW			
Grammar Checker Score				Grammar Checker Score			
LLD	predicted score <= 2	predicted score > 2	total		predicted score <= 2	predicted score > 2	total
mean score <= 2	0 (0%)	131 (14%)	131 (14%)	mean score <= 2	54 (6%)	77 (8%)	131 (14%)
mean score > 2	0 (0%)	805 (86%)	805 (86%)	mean score > 2	20 (2%)	785 (84%)	805 (86%)
total	0 (0%)	936 (100%)	936 (100%)	total	74 (8%)	862 (92%)	936 (100%)
% correctly predicted			86%	% correctly predicted			90%
% of score <= 2 correctly predicted			0%	% of score <= 2 correctly predicted			41%
% of score > 2 correctly predicted			100%	% of score > 2 correctly predicted			98%
Grammar Checker/Score: PE >=5				Grammar Checker/Score: RW >= 5			
HLD	predicted score >= 5	predicted score < 5	total		predicted score >= 5	predicted score < 5	total
mean score >= 5	23 (2%)	199 (21%)	222 (24%)	mean score >= 5	58 (6%)	164 (18%)	222 (24%)
mean score < 5	12 (1%)	702 (75%)	714 (76%)	mean score < 5	32 (3%)	682 (73%)	714 (76%)
total	35 (4%)	901 (96%)	936 (100%)	total	90 (10%)	846 (90%)	936 (100%)
% correctly predicted			77%	% correctly predicted			79%
% of score >= 5 correctly predicted			10%	% of score >= 5 correctly predicted			26%
% of score < 5 correctly predicted			98%	% of score < 5 correctly predicted			96%

Table 10a - Summary of scoring performance showing accurate scoring for LLD and HLD decisions for combined sample as model and combined sample as data (model: combined; sample: combined)

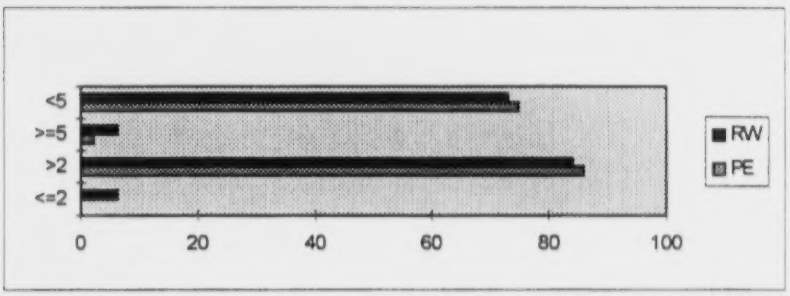


Table 10 presents a summary comparison of the two best grammar-checking programs on the combined samples. When the two samples are combined, RW shows clearly superior agreement for the lld decision. While the **overall** percentage agreement favored RW by only 4% (90% vs. 86%), PE did not classify any papers at the 2 or below level. Of the 131 papers that the raters classified as 2 or lower, RW agreed on 41%. However, RW also placed 20 (about 2%) of the "true" greater than 2 papers in the 2 or less category.

Inspection of the lower section of Table 10 (the results for the hld in the combined sample) shows a relatively equivalent **overall** agreement rate with 83% for RW and 82% for PE. PE does somewhat better than RW in predicting the hld classification but also makes more errors than RW in placing essays in the high group which belong in the remaining group.

Table 11 presents a summary of the types of errors that were made by the two software packages.

Table 11 - Summary of Errors in Prediction by Error Type		
Method	**lld decision**	**hld decision**
PE	pred(high | true low) = 100%	pred(high | true low) = 5%
RW	pred(high | true low) = 59%	pred(high | true low) = 3%
PE	pred(low | true high) = 0%	pred(low | true high) = 69%
RW	pred(low | true high) = 3%	pred(low | true high) = 76%

The percentages in Table 11 suggest that the clear difference between the two procedures is with respect to the lld decision. As indicated earlier, RW seems to be superior here. Inspection of the types of errors involved in the hld decision suggests little difference between the grammar checking programs. The one exception to this might be if predicting that a paper is less than 5 when it is a "true" 5 or greater is considered a serious mistake, i.e., would have serious consequences. If that were the case, PE might be considered for hld decisions.

Table 12 presents the significant predictors from the logistic regressions for the two grammar-checking programs.

Table 12 - Logistic Regression Weights For the Various Models and Decisions			
lld			
Predictors	**PE Model r-biserial = .629**		
	Reg. Wt	**Std. Error**	**t Stat.**
elegance	.110	,015	7.30
emphasis	.377	.051	7.40
grammar	-.042	.036	-1.17
	RW Model r-biserial = .896		
RWcon	.472	.100	4.73
discourse	.319	.063	5.07
elegance	.377	.077	4.89
grammar	.378	.009	4.24
hld			
Predictors	**PE Model r-biserial = .557**		
	Reg. Wt	**Std. Error**	**t Stat.**
elegance	-.012	.018	-.67
emphasis	-.240	.068	-3.53
grammar	.017	.040	.44
	RW Model r-biserial = .564		
RWcon	-.160	.043	-3.70
discourse	-.134	.033	-4.06
elegance	-.123	.036	-3.41
grammar	-.241	.042	-5.80

Inspection of Table 12 indicates that for the lld decision only elegance and emphasis were statistically significant ($|t| > 2$) in the PE model. The RW lld decision model had four significant predictors: consistency, discourse, elegance, and grammar. The r-biserial shown on the model line is a single index of the relationship between the predicted classification and the actual classification. As one might expect, the r-biserial for the RW model is considerably higher than that for the PE model for the lld decision.

Table 12 indicates that each model used the same predictors for the lld decision and the hld decision. Only the signs changed because the coding of

the hld decision was the reverse of that of the lld decision. Within models the pattern of the significant regression weights is similar, suggesting that the weighting function just "shifted up" from the lld decision to the hld decision. The r-biserials are almost the same for the hld decision, suggesting there is little difference between the two models for the hld case.

Linguistic Analysis

Scores estimated by RW were correctly predicted for 26.8 % of the high scoring (>=5) and 35.6% of the low scoring (=<2) essays, as compared with scores assigned by human graders. These results show that RW was able to estimate scores for approximately one-third of the essays in this study. Though this is a promising result, we believed that a review of the essays which were incorrectly scored[6] by RW would provide information as to how RW's performance could be improved. With regard to this, we addressed the following two questions: a) Overall, why did RW correctly predict more low scoring essays than high scoring ones? and b) How can the overall percentage of essays correctly scored by RW be increased?

Linguistic Analysis - Method and Discussion

[6]These were the essays scored by RightWriter which were assigned a score of 5 or greater as compared to a score of 1 or 2 by human graders, and, conversely, where a score of 2 or less was assigned to essays given a score of 5 or 6 by human graders.

We initially extracted a total of 40 essays, 10 from each of the four prediction groups shown in Table 13. Our intention was to do a preliminary linguistic analysis to see how specific linguistic features were evaluated by RW.

Table 13 - Four Prediction Groups (high = >=5 and low = <=2)	
cor2	correctly predicted low
cor5	correctly predicted high
incor2	incorrectly predicted high
incor5	incarcerate predicted low

We examined each essay, along with the *error categories* carried over from the grammar-checking program comparison. We observed that the high- and the low-scoring essays (independent of whether they were accurately predicted by the grammar checker or not) differed with regard to the overall number of errors reported. The number of errors was higher for high-scoring ("good") essays than for low-scoring ("poor") essays. Incorrectly predicted high-scoring essays (incor5) had fewer errors than correctly predicted ones (cor5), and incorrectly predicted low-scoring (incor2) essays had more errors than the truly low-scoring ones (cor2).

We observed that RW reported significantly more errors for the "good" (high-scoring) essays, and fewer errors, or even absence of errors for the low-scoring essays. Since grammar-checking programs presuppose a certain competence level on the part of the writer, this inverse relationship was unexpected. Still, the total absence of any reported errors in the face of obvious violations of English grammar in a few of the essays needs to be

examined[7]. Furthermore, the overall number of errors per essay is too gross a measure, as it does not take into account the varying lengths of the essays: "good" essays were also longer essays than "poor" ones, a correlation that has been established elsewhere (see Breland, et al (1987) and Breland et al (1994). A comparison of the essays with respect to their errors per essay-length ratio did not yield any drastic differences among the various groups of the sample.

The initial category analysis provided us with little information about the linguistic differences between the essays in the four prediction groups. We concluded that although the category analysis was useful as a mapping device over the four grammar checkers, it appeared to be too general for the purposes of a finer-grained analysis of RW performance. The actual error classes generated by RW proved to be more informative. We extracted RW's error analysis of the essays by hand. We were able to do this analysis on a total of 20 of the essays, 5 for each of the four prediction groups.

Even for this small set of essays, when we used the RW error classes, we were able to find some associations between general linguistic information picked up by RW and its score estimations. Specifically, all essays in which RW estimated a high score (cor5 and incor5), and also some essays of the incor2 group, were critiqued for excessively long sentences or

[7] see Bowyer (1989) for a detailed discussion of RW's procedures for analyzing grammatical errors.

paragraphs.[8] Cor5 essays had the highest occurrence of this error class. Cor5 and incor5 contained a considerable number of passive constructions according to RW. Essays that were incorrectly predicted to have high scores (incor2) also had more passive constructions than the essays given a low score by human graders. Usage errors[9] were reported for high-scoring essays but were more or less absent in the low-scoring ones.

The overall length of the essays scored incorrectly by RW were, on the average, longer than the "poor" essays and shorter than the "good" ones. With regard to the number of style, grammar, and usage errors, the number of errors generated for incorrectly-scored essays was in between the truly good and the truly poor essays. As indicated before, the ratio of a given error type and the overall length of the essay might provide a more informative measure than numbers alone. A larger sample might show additional variables, or statistically more significant variables, for automatic-scoring procedures.

We observed some general linguistic features distinguishing high- and low- scored essays which RW did not appear to pick up. In general, the high scoring essays had better syntax, vocabulary, style, and organization than the low scoring ones. Their sentences were not only longer, but often more

---

[8] RW reported this as "excessively long sentence," with a threshold of 25 words per sentence, often followed by a suggestion to split the sentence in two.

[9] RW's categories for usage errors include vagueness, wordiness, redundancy, use of slang, and technical jargon.

complex, with proper conjunctions and, or and complementizers that, why. The low scoring essays had shorter and also incomplete sentences. Complex sentences often lacked sentence connectives (e.g., in addition, furthermore). These features are illustrated in the high-scoring and low-scoring essays below.

High-Scoring Essay (COR5)

Whether newspapers are better sources of news than radio or television depends on each person's perspective or point of view. Personally, I prefer newspapers to any other source of information.

Most newspapers give a complete and explanatory report on every day news. Each issue is considered and discussed in a clear and impertial way, this is very important so that the news don't depend on the writer's perspective.

Moreover, unlike television or radio in which the information is given in a specific moment and is not repeated later, newspapers give the reader the chance to read again the information and even keep it for after use.

In addition, news broadcasted in television and radio tend to have less or more importance according to the way they are broadcasted by the journalist. If the reporter agrees on the topic that is being discussed he would probably tend to emphasize the information, also if he doesn't agree, the importance of the report will probably decrease.

Newspapers are not only less personalized than television and radio but they are also more precise and complete. Most of the times they include graphs, statistics, opinions and pictures that help the reader get a clearer idea of the situation that surrounds a certain issue.

To sum up, newspapers have all the conditions that are necessary in order to have good information. That is: they are neutral, precise and give a complete account of the news regardless the writer's personal opinion or political point of view. These are the main reasons why I prefer newspapers to any other source of information.

Low-Scoring Essays (COR2):

I think the TV is very good to follow the news because the TV is follow the news in live time and get the correct new to people.

Some other general characteristics of the essays pertaining to content rather than surface syntax distinguished the "good" and the "poor" essays.

For instance, high-scoring essays logically presented opinions by providing ever stronger pros and cons to support them - features that are impoverished or altogether absent in the low-scoring ones.

Discussion

In Tables 6 through 10 and their related analyses, there are two fundamental questions that we sought to answer. The first of these is whether we could construct a model based on the output of grammar-checking programs that could predict the score a human rater would assign to a TWE essay. Part of this question includes what the formulation of the model would be, and part is what sort of accuracy could be attained with such a model. Of the fifteen variables derived from the grammar-checking programs' error messages, only those categorized as concision, discourse, elegance, and grammar were significant in predicting essay scores.

The best-performing grammar-checking programs were RW and PE. The analysis of these two grammar-checking programs proved to be highly correlated with being able to predict the scores of certain essays. The outcome that RW was the superior performer in the lld decision ran counter to our intuition. As mentioned early in this report, because PE uses a more sophisticated and perhaps more well-founded approach to analysis, we believed it would outperform all of the other grammar-checking programs in its ability to recognize and classify errors in writing. This was not the case.

This outcome might be explained in terms of RW's ability to identify patterns in writing. If the patterns incorporated into RW were such that a) they encompassed a wide variety of writing phenomena and b) they could be applied with a high degree of accuracy, then RW could possibly perform better than PowerEdit as was the case in our analysis. An interesting question to explore is the accuracy with which these grammar-checking programs assign errors to samples of writing. If we had some idea of the actual error rate, this might give us a better way to estimate the performance of a particular grammar-checking program.

At the outset we need to know what we can expect from a scoring model based on grammar-checking programs. To answer this question, three summary tables have been prepared. Tables 14 through 16 summarize the scoring performance of the models.

Table 14 shows, for RW and PE, the total number of essays for which a score was correctly computed. This table represents the combined scoring performance for all models and for all scoring categorizations. The bottom line of the table indicates that, overall, for placing essays into the >=5 category and the <=2 category, PE correctly placed essays 12% of the time and RW correctly placed essays 31% of the time. This essentially tells us that we could expect RW to classify correctly, overall, about 1/3 of the essays that would have to be scored, leaving the remaining 2/3's of the essays for human raters.

Table 14 - Overall comparison of score predication performance					
	PE	**RW**	**Score Prediction**	**Model**	**Data**
	0	42	<=2	1	1
	0	49	<=2	2	2
	0	16	<=2	1	2
	0	30	<=2	2	1
	0	41	<=2	1+2	1+2
	20	27	>=5	1	1
	35	27	>=5	2	2
	24	30	>=5	1	2
	31	24	>=5	2	1
	10	26	>=5	1+2	1+2
Average % computed correct overall	**12**	**31.2**			

When we consider individually how the models performed overall we see that in the case of the >=5 categorization, performance of each of the

grammar-checking programs was about the same, yielding a correct scoring categorization of about 25% overall.

Table 15 - Scoring performance for essays scored >=5					
	PE	**RW**	**Score Prediction**	**Model**	**Data**
	20	27	>=5	1	1
	35	27	>=5	2	2
	24	30	>=5	1	2
	31	24	>=5	2	1
	10	26	>=5	1+2	1+2
Average % computed correct overall	**24**	**26.8**			

Likewise, considering scoring performance for the <=2 categorization decision shows us that we could expect RW to correctly categorize 35% of the essays processed - again roughly 1/3 of the essays. In an essay population of 800,000 essays where approximately 10% would be rated score <= 2, this scoring procedure would result in 26,000 essays not having to be examined by human raters. Over the whole sample of essays this represents about 3% of the essays. Clearly the scoring procedure would have to be improved if we were to adopt it as part of the process of scoring TWE essays.

One important consideration for using this model is how to tell when the procedure produces a true or false score. In other words, one of the important aspects of this model is that we are sure 35% we know were placed in the <=2 score category, were correctly placed. We know this because, associated with each score estimation is the probability that the essay should be assigned to a category. By comparing the magnitudes of the probabilities

we can accurately select the essay score category. We can use the difference in magnitude to create an estimate of the reliability of assignment to a score category.

Table 16 - Scoring performance for essays scored <=2

	PE	RW	Score Prediction	Model	Data
	0	42	<=2	1	1
	0	49	<=2	2	2
	0	16	<=2	1	2
	0	30	<=2	2	1
	0	41	<=2	1+2	1+2
Average % computed correct overall	0	35.6			

From the linguistic point of view, if surface criteria such as essay length, number of words per sentence and number of words per paragraph are fairly reliable indicators of the writing skills of a non-native speaker of English , and if a proliferation of passive constructions in an essay is another measure of competence, then RW could be an aid in estimating scores of essay items. Enlarging the pool of correctly-scored essays by RW could be achieved by lowering or raising the error threshold for the variables indicated. A larger sample should be studied for this purpose and might show possible correlations with other error types. For instance, with regard to the latter, wordiness or the use of clichés presupposes a greater competence of English and might go hand-in-hand with essay length as an indicator for a high-scoring essay.

It would be beneficial to re-run this analysis, using RW error classes instead of the categorizations created for the initial study. In a second pass, we might find that RW is able to be a more efficient score estimator if its fine-grained set of categories is used as variables. A more thorough analysis might enable us to collapse certain categories, eliminate others, add categories, and identify additional factors which would help improve RW's performance in this task.

Conclusions

In this study we have investigated how well one automated model of scoring can predict expert ratings of essays produced as part of the TWE. This model is based on using a commercial grammar-checking program to analyze an essay, the categorization of the messages produced by the program analysis, and the application of a statistical model to predict the score for an essay based on the cumulative summary of errors categories.

Our results showed that: 1) a model could be constructed using the output of commercial grammar-checking programs; 2) approximately 30% of essays analyzed could be scored correctly; 3) the scores derived from the scoring model could be accepted as accurate; and 4) the number of essays scored by this procedure does not yet warrant its application in a practical setting.

This latter aspect of the study indicates that more research would be required to determine whether such a model could effectively score 50%, 60%,

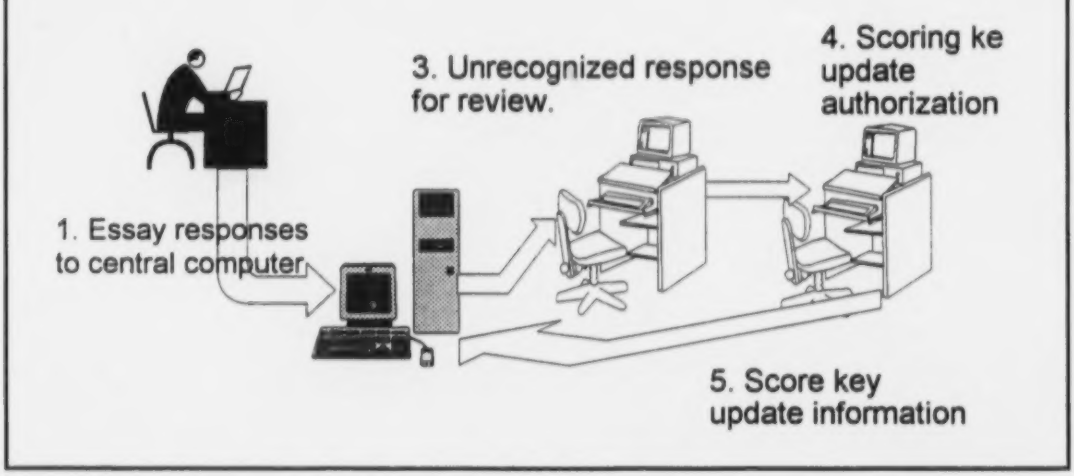
or even 90% of the essays. As suggested by the linguistic analysis, it is entirely possible that the need to create cumulative summaries of the error messages produced by a grammar checker could have obscured the characteristics of an essay to such an extent that any model constructed would not be sufficiently accurate to estimate many of the essay scores. A potential next step for this work would be to analyze the essays and create a finer grained analysis of the kinds of errors that appear on different essays. Having done this, we would use this information to construct a new model. This model could then be evaluated in a manner similar to the presented approach.

In general, we have consistently viewed the process of scoring complex constructed responses as a multi-level process. At different levels of the analysis, different procedures might be appropriate. An advantage to the approach described in this report is that it rapidly obtains an estimated essay score; more sophisticated approaches would require more analysis time. The model-based approach might be best as the first level of a complex scoring procedure. Further investigation is needed to determine if this procedure functions well as a part of a more complex scoring procedure.

Another possibility for investigation is the overlap between the two decision sets. In other words, we did not examine the essays in the 2-5 range as scored by the lld and hld scorings. The essays contained in this overlap set might in fact constitute another viable scoring group.

One last consideration is how the scoring procedure described in this study would be integrated into an operational setting. Given that ongoing development into this scoring process yields more effective scoring results, such a procedure may be integrated in a computer-assisted scoring model. In this model, a computer system scores essays using a procedure like the one described. In the event that the system cannot score an essay, the essay is sent to a human rater for scoring. Automatic scoring of other essays continues while a human rater scores the essay that could not be scored by the scoring system. When a score has been assigned, the rater will send the scored essay back to the scoring system. The system will integrate the scored essay into its database of scored essays and modify its scoring rubric appropriately if indicated by the human rater. Figure 5 depicts one possible operational setting for scoring essays.

Figure 5 - Operational setting for essay scoring


3. Unrecognized response for review.
4. Scoring ke update authorization
1. Essay responses to central computer
5. Score key update information

References

Bowyer, J.W. (1989). A Comparative Study of three Writing Analysis Programs. Literary and Linguistic Computing, vol.4., no.2.

Breland, H.M., Camp, R., Jones, R.J., Rock, D., & M. Morris (1987). Assessing Writing Skill. New York: College Board.

Breland, H.M., Danos, D.O., Kahn, H.D., Kubota, M.Y., & Bonner, M.W. (1994). Performance versus objective and gender. Journal of Educational Measurement, vol. 31.

Frase, L.T., & Faletti, J. *Computer Analysis of the TOEFL Test of Written English*. Proposal submitted to the TOEFL Research Committee. Princeton: April 1991.

Page, Ellis B. (1966). The Imminence of Grading Essays by Computer. Phi Delta Kappan. January, 238 - 43.

Page, Ellis B. and Petersen, N. (1995). The Computer Moves Into Essay Grading; Updating the Ancient Test. Phi Delta Kappan. March, 561-65.

Appendix A
Sample Grammar Check Outputs

A.1 Correct Grammar

Correct Grammar's output consisted of two parts, a summary and a detailed list of diagnostic messages embedded in the essay. A partial sample of output is shown below:

7	paragraphs, average	2.4	sentences each
17	sentences, average	16.3	words each
278	words, average	4.7	letters each
156	syllables per 100 words		
3	passive sentences	17 %	of total
1	long sentences	5 %	of total
2	misspelled words	99 %	correct
7	other errors corrected	58 %	correct
1	sentences hard to read	94 %	correct
Flesch Reading Ease score		58.3	Fairly Easy
Grade level required		9	
U.S. adults who can understand		85 %	
Flesch-Kincaid grade level		9.1	
Gunning Fog Index		8.3	

... [-- Sentence exceeds recommended length. --] I remember the times when our science teacher took us outdoors on nature trips
Opening up a whole new world, if we had
only read about what a flower or a bird or
an animal was, but never [-- Overused modifier. Use sparingly. --] actually saw one,
I am sure that I would not retain such
wonderful memories. ...

A.2 Grammatik

Grammatik also contained individual diagnostic messages and summary information. A partial sample of Grammatik's output is shown below:

Check: each and every

Problem: Hackneyed, Clichè, or Trite

Advice: Try `each` or `every`.

Check: is handled

Problem: Passive voice

Advice: Passive voice: `is handled`. Consider revising using active voice.

========== Grammatik III - Version 1.02 ==========

Summary for \grammar\essays\file1

Problems marked/detected: 13/13

Readability Statistics

Flesch Reading Ease: 59
Gunning's Fog Index: 11
Flesch-Kincaid Grade Level: 9

Paragraph Statistics

Number of paragraphs: 1
Average length: 17.0 sentences

Sentence Statistics

Number of sentences: 17
Average length: 16.3 words
End with '?': 0
End with '!': 0
Passive voice: 2
Short (< 14 words): 9
Long (> 30 words): 1

Word Statistics

Number of words: 278
Prepositions: 17
Average length: 4.71 letters
Syllables per word: 1.55

A.3 RightWriter

RightWriter also contained individual diagnostic messages embedded in the text and summary information. A partial sample of Grammatik's output is shown below:

```
    Nowadays, schooling becomes a complusory performance in one's life.
Everybody will definitely go to school once in their lives.  However, some
        <<* U9. IS THIS JUSTIFIED? definitely *>>
people are afraid of going to school because they are scared by the toughness
              <<* S1. PASSIVE VOICE: are scared *>>^
and the demand of their teachers.  The students find their teachers boring and
        <<* S4. IS SENTENCE TOO DIFFICULT? *>>
so they lose their interest in exploring the knowledge. ...

              <<** SUMMARY **>>

  The document file1 was analyzed using the rules for
  General Business writing at the General Public
  education level.  It is a Standard ASCII document.
  The marked-up copy is stored in the file FILE1.OUT.

READABILITY INDEX:  9.92

  4th     6th     8th     10th    12th    14th
   |****|****|****|****|****|****|   |   |   |   |
  SIMPLE    | ------ GOOD ------ |          COMPLEX
  Readers need a 10th grade level of education.


STRENGTH INDEX:  0.43

  0.0             0.5             1.0
   |****|****|****|****|*  |   |   |   |   |   |
  WEAK                          STRONG
  The writing can be made more direct by using:
              - the active voice
              - shorter sentences
              - fewer weak phrases
              - more common words


DESCRIPTIVE INDEX:  0.49

  0.1         0.5         0.9     1.1
   |****|****|****|***|   |   |   |   |   |   |
  TERSE  | ------------ NORMAL ------------ |    WORDY
  The use of adjectives and adverbs is normal.


JARGON INDEX:  0.23
```

A.4 PowerEdit

PowerEdit took the sentences one by one and gave individual diagnostic messages. A partial sample of PowerEdit's output is shown below:

Sentence # 6 of 8

On the other hand, if students do not like learning, their

countries will suffered many problem.

[286/1] <Gram> "Will" and "suffered" do not seem to belong together. Should one be removed? Has a word been left out?

[53/3] <Usag> "Many" does not seem to match "problem." Do they belong together? Are they part of a special phrase? Has a word such as "that" been deleted? Is there a missing comma?

[59/1] <Tran> Is "on the otherhand, if students do not like learing" the introductory part of this sentence? If so, the introduction may be too long for this sentence. You may want to re-organize this sentence.

[222/1] <Logc> The words "like learing" may be used incorrectly, or the following words may be unclear.

[221/12] <Logc> Could "on the" be worded a little more clearly?

[221/9] <Logc> Be careful with "like learing" and the surrounding words. This wording may be difficult to understand or part of a special phrase.

[172/1] <Eleg> "Learing" has a literary sound to it.

Appendix B
Categorization of Errors from Grammar Checkers

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Balance	288		This sentence might read better if the subject were shorter in relation to the predicate. Try to make the predicate longer than the subject by putting any new information in it or by reducing the old information in the subject.			
Cohesion	014	Grammar/ Subjects	The subject for "are" may not be apparent or may be missing. Can you clarify "in the other way?"			
Cohesion	065	Style/ Writing Style/ Redundant Subjects		29. These words may be redundant; consider omitting them. 30. Redundant expression. Use ... instead.	26. Redundant phrase	S14. Consider omitting: ... U13. Redundant: ... U13. Redundant. Replace ... by ...
Cohesion	220	Grammar/ Modification/ Non-Essential	If the phrase "because ... has a strong link with the environment and exposure to nature" is not essential to the sentence, it may need some punctuation around it.			
Cohesion	229	Style/ Word Selection/ Afterthought	A sentence beginning with "in addition" seems like an afterthought. You may want a stronger introductory word or phrase.			
Cohesion	240	Clarity/ Ambiguity Clarity/ Insufficient Information	"Being from" and the following words may be unclear to some readers. Should they be rewritten?			S15. Is this ambiguous: ...
Cohesion	268	Clarity/ Insufficient Information				
Concision	066	Grammar/ Usage/ Incorrect	"What" and the following words may be difficult to understand. Can you clarify this sentence? Are there special phrases in this sentence?			G9. Is ... being used correctly G12. Is ... correct S4. Is Sentence too difficult
Concision	124	Style/ Word Selection/ Wordy	"First of all" may be considered wordy.	26. Wordy expression. Consider ... instead.	18. Long-winded or wordy 36. Longwinded or wordy	U11. Wordy: ... U12. Wordy: Replace ... by ...
Concision	137	Style/ Writing Style/ Redundant Subjects	"Each individual" is redundant. Could the same point be made without repetition?	29. These words may be redundant; consider omitting them. 30. Redundant expression. Use ... instead.	26. Redundant phrase	S14. Consider omitting: ... U13. Redundant: ... U13. Redundant. Replace ... by ...

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Concision	138	Tone/ Complexity/ General				S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Concision	162	Style/ Word Selection/ General	"As a result" may be replaced by a single word.			G12. Wrong word. Replace ... by ... S22. Should ... be ...
Concision	178	Tone/ General/ Necessary	"Etc" may not be needed to convey your idea.			
Concision	207	Clarity/ Wordiness/ Redundancies	"Literally" and "right" may say nearly the same thing twice. Make sure that your meaning is clearly expressed.			
Concision	416	Clarity/ Readability/ Difficulty	The words around "clear" may be overly complex. Can you clarify this sentence?			S4. Is Sentence too difficult S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Concision				7. Consider deleting the repeated word ... . 40. Consider changing or deleting ... . 46. Consider deleting		G13. Repeated word.
Discourse	007	Clarity/ Theme	You may need to strengthen the main topic and focus of this sentence.			
Discourse	015	Grammar/ Subjects	The main idea in this sentence may be unclear. Could you clarify?			
Discourse	017		The clause "depending these three graphs shown" may be difficult to read. A verb seems to be missing or very weak, and may cause ambiguities.			G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S4. Is Sentence too difficult S5. Use verb form. Replace ... by ... S17. Weak: ... S18. Weak: Replace ... by ...

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Discourse	018	Clarity/ Theme	The main action in this sentence may not be clear. Is there a verb or some punctuation missing? Is this sentence a fragment?	2. This does not seem to be a complete sentence. 13. This sentence does not seem to contain a main clause.	30. Incomplete sentence	G2. Is this a complete sentence P3. Incomplete sentence or missing comma
Discourse	019	Grammar/ Subjects	The subject in this sentence may be unclear. Is it missing? Is this sentence a fragment? Is there a comma missing after the introductory part of the sentence?	2. This does not seem to be a complete sentence. 13. This sentence does not seem to contain a main clause.	30. Incomplete sentence	G2. Is this a complete sentence P3. Incomplete sentence or missing comma P3. Is comma missing after ...
Discourse	044	Clarity/ Theme	The main action in "makes learning enjoyable he would help the people" may be unclear; does this sentence mean what you want it to, or should something be added or left out?			
Discourse	067	Clarity/ Readability/ Difficulty	This sentence may be difficult to understand. Is this a sentence fragment? Should you consider rewriting? Check the sentence around "Farms."	2. This does not seem to be a complete sentence. 13. This sentence does not seem to contain a main clause.	30. Incomplete sentence	G2. Is this a complete sentence S4. Is Sentence too difficult P3. Incomplete sentence ormissing comma
Discourse	143	Tone/ General/ Legalese	"So as" is specific to legal audiences.			U7. Legalese: ...
Discourse	181	Style/ Sentence Length	This sentence may be too long and too complex for your reader. Can you shorten or clarify it?	10. Sentence exceeds recommended length.	17. Long sentence	S3. Long Sentence: ... S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Discourse	191	Clarity/ Clarity/ Usage Related				
Discourse					19. Paragraph problem	
Discourse					47. One sentence paragraph	
Discourse						S6. Long Paragraph: ...
Elegance	106	Tone/ Complexity/ Alternative Wording		39. Consider rewriting the awkward expression ...	45. Clumsy or awkward	S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Elegance	111	Style/ Word Position/ Initial Wording	You could replace "characters whose behavior" with "characters the behavior of which" or with some version of this.			G12. Wrong word. Replace ... by ... S22. Should ... be ...

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Elegance	112	Tone/ Formality/ General		68, 76, 77. Avoid using contractions like ... in formal writing		S21. Contraction
Elegance	115	Tone/ Idiomatic/ Slang		42. Nonstandard. Consider ... instead. 56. rewriting the nonstandard compound		
Elegance	116	Tone/ Derogatory/ Vulgar	"Pissed" may be considered vulgar by some audiences.			U17. Offensive: ...
Elegance	117	Tone/ Derogatory/ Obscene				U17. Offensive: ...
Elegance	118	Tone/ Derogatory/ Obscene				U17. Offensive: ...
Elegance	119	Tone/ Formality/ General	"Anyway" may be too informal for some audiences.	35. Informal. Use ... or ... unless referring to something like grapes. 50. Colloquial modifier. 71. Invalid contraction used in sentence.	31. Informal or colloquial 35. Informal or illiterate	S21. Contraction U1. Colloquial: ... U2. Colloquial. Replace ... by ...
Elegance	126	Clarity/ Nominalizations	Words like "chosen" following weak verbs like "have" should be avoided. Try to put the action expressed in "chosen" into a verb form that replaces "have."			G4. Wrong verb, replace ... by ... G8 Is ... the correct form of the verb S5. Use verb form. Replace ... by ... S17. Weak: ... S18. Weak: Replace ... by ...
Elegance	134	Tone/ General/ Onomatopoeia				
Elegance	135	Tone/ Idiomatic/ Cliché		69. Use ... to specify the topic; ... indicates date or location.	16. Hackneyed, Cliché, or Trite	S16. Cliché: ...
Elegance	136	Tone/ Formality/ General				
Elegance	139	Tone/ General/ Religious				
Elegance	144	Tone/ Idiomatic/ Jargon	"And/or" is specific to certain audiences, and should be used carefully.	43. Avoid jargon words like ... . 52. Jargon.	24. Jargon	S8. Computer jargon: ...
Elegance	145	Tone/ Vagueness/ Acronym				
Elegance	146	Tone/ General/ Foreign	"Favour" is a foreign language expression.		22. Foreign	
Elegance	147	Tone/ Idiomatic/ Folksy				
Elegance	150	Tone/ General/ Overused	"A lot" tends to be overused. Could you use a word that is more specific or descriptive?	18. Overused modifier. Use sparingly. 19. Overused. Use sparingly.		S19. Overused: ...

BEST COPY AVAILABLE

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Elegance	152	Tone/ Derogatory/ Sexist	"Mankind" may be considered offensive by some audiences. You may want to use a word that does not specify gender.		23. Gender Specific 33. Gender Specific	U5. Is this sexist? ...
Elegance	153	Tone/ Emphasis/ Sensationalism				
Elegance	154	Tone/ Vagueness/ Abbreviation	"Etc" is an abbreviation and may be inappropriate for formal writing.	36. The abbreviation ... should be spelled out in formal writing.		
Elegance	157	Tone/ Emphasis/ General	"Actually" is emphatic and should be used carefully.			
Elegance	165	Tone/ Derogatory/ General			34. Negative usage	S11. Is sentence too negative U17. Offensive: ... U21. Negative: ...
Elegance	176	Tone/ Derogatory/ General				U17. Offensive: ...
Elegance	197	Clarity/ Complex/ General Relationships				S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Elegance	213	Clarity/ Nominalizations	The word choice in "with this consideration in mind we have to observe that what may be bad or outrageous behavior for some, its common behavior for others" keeps the reader at a distance from the action or process.			
Elegance	241		(Choppy Flow) This sentence consists of many small parts. The essential parts may be difficult to find. Can you clarify?			S4. Is Sentence too difficult
Elegance	287	Tone/ Idiomatic/ Euphemism				U20. Misleading euphemism: ...
Elegance					13. Number Style	
Elegance					41. Overstated or pretentious	
Elegance						S7. Sentence Begins with but S8. Sentence Begins with conjunction
Emphasis	033	Style/ Passive Voice		20. This main clause may contain a verb in the passive voice.	20. Passive voice	S1. Passive voice ...
Emphasis	179	Style/ Word Position/ General				

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Emphasis	196	Clarity/ Readability/ Position	This sentence may be more understandable if the word "simply" were moved toward the end of the sentence.			
Emphasis	219	Clarity/ Readability/ Position	Make sure that "is" should end this sentence.			
Grammar	001	Grammar/ Agreement/ Subject-Verb	The subject for "are not" may be unclear. If it is "some," then "are not" must agree in number with it. The structure of this sentence may need to be clarified.	8. The word ... does not agree with ... . 15. The verb after ... must agree in number with the following noun phrase. 59. agrees with the subject	7. Verb agreement 38. Number agreement	G1. Do subject and verb agree in number
Grammar	002	Grammar/ Agreement/ Verb-Complement	"Changes" may be the wrong word. Should it agree in number with "is"? Is it part of a special phrase?	8. The word ... does not agree with ... . 15. The verb after ... must agree in number with the following noun phrase.	7. Verb agreement 38. Number agreement	
Grammar	011	Grammar/ Usage/ Determiners	"A" may be inappropriate with "statements." Should it be deleted? If not, the words between "A" and "statements" may be overly complex, may be part of a special phrase, or may have some important words deleted.			G6. Replace A by AN G7. Replace AN by A
Grammar	030	Grammar/ Verbs/ Usage		65. Consider using a form of ... with ... or replacing with .... or ...		G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S5. Use verb form. Replace ... by ...
Grammar	041	Grammar/ Coordination	This sentence may be too complex. The words around "will be" and "living" may be difficult to understand. Are the verb tenses consistent? Could you clarify?			G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S4. Is Sentence too difficult S5. Use verb form. Replace ... by ... S12. Can simpler terms be used S13. Replace ... by simpler .... S13. Replace ... form of simpler ...

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Grammar	045	Grammar/ Missing Words	This sentence may be difficult to read around "they." Is there a verb missing, or is the sentence structure improperly coordinated or overly complex?			
Grammar	047	Grammar/ Verbs/ Order	"Are" and "may" appear to be two verbs in the same phrase. "Are" may need to be the first verb in the phrase. Are these words used correctly? Is there a comma missing somewhere? The words between "Are" and "may" may be overly complex.			G9. Is ... being used correctly G11. Is ... correct P3. Is comma missing after ...
Grammar	049	Grammar/ Modification/ Incorrect	"Common" cannot usually have modifying words such as "one" in front of it.			G9. Is ... being used correctly G11. Is ... correct
Grammar	050	Grammar/ Usage/ Determiners	"Atmosphere" may need a word such as "the," "a," "an," "some" in front of it, or may be part of a special phrase.		3. Article usage	G6. Replace A by AN G7. Replace AN by A
Grammar	051	Grammar/ Verbs/ Forms				G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S5. Use verb form. Replace ... by ...
Grammar	052	Grammar/ Verbs/ Forms				G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S5. Use verb form. Replace ... by ...
Grammar	079	Grammar/ Fragments		2. This does not seem to be a complete sentence. 13. This sentence does not seem to contain a main clause.	30. Incomplete sentence	G2. Is this a complete sentence P3. Incomplete sentence or missing comma
Grammar	094	Grammar/ Usage/ Inappropriate	"Right" may be inappropriate with "through." Is "Right" modifying "through"? If so, it may not be properly used, or may be redundant.	29. These words may be redundant; consider omitting them. 30. Redundant expression. Use ... instead.	26. Redundant phrase	S14. Consider omitting: ... U13. Redundant: ... U13. Redundant. Replace ... by ...
Grammar	200	Clarity/ Clarity/ Indirect Questions				P5. Quotations introduced by that are indirect
Grammar	202	Clarity/ Ambiguity Clarity/ Clarity/ Negations				S15. IS this ambiguous: ...

BEST COPY AVAILABLE

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Grammar	211	Clarity/ Wordiness/ Run-on/ Fused	The sequence of words "farmers just used animals" may be incorrect. A comma, hyphen or a subordinator such as "that" may be needed. Can you clarify?	25. This appears to be a run-on sentence.		G3. Split into 2 sentences
Grammar	215	Clarity/ Wordiness/ Run-on/ Fused	This sentence may run through several ideas. Should the ideas be more clearly seperated?	25. This appears to be a run-on sentence.		G3. Split into 2 sentences
Grammar	225	Grammar/ Major/ Comma	The comma after "decrease" could be removed. Make sure that you are consistent with your punctuation before conjunctions.			
Grammar	257	Style/ Passive Voice	There is more than one passive verb like "be broken" in this sentence. There may be a more direct way to state the actions in this sentence. See 'Tutorial' for a detailed explanation.	20. This main clause may contain a verb in the passive voice.	20. Passive voice	S1. Passive voice: ...
Grammar	259	Clarity/ Readability/ Difficulty				S4. Is Sentence too difficult
Grammar	276	Clarity/ Complex/ General Relation	Are the words "this kind of teachers" part of the same phrase? If so, they should agree in number. If not, then they may be unclear to the reader or part of a special phrase.	8. The word ... does not agree with	38. Number agreement	
Grammar	286	Grammar/ Usage/ General Relation	"Will" and "depends" do not seem to belong together. Should one be removed? Has a word been left out?			
Logic	003	Clarity/ Readability/ Flow	This sentence does not flow well. "to they ... it" starts the area of poor flow. Is "to they ... it" used correctly? Can you clarify?			G9. Is ... being used correctly G11. Is ... correct
Logic	004	Clarity/ Sprawl	"Farms and farm population" may be difficult to read or may contain too much information or a side comment. Could it be clarified?			S4. Is Sentence too difficult
Logic	012	Grammar/ Insufficient Information	"Teaching interesting" may be difficult to read. Does it need a comma? Should it be rewritten? Are there an implied subject and verb?			

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Logic	013	Clarity/ Wordiness/ Introductions	The part of this sentence starting with "otherwise we may bring disasters, such as" and ending with "war and force, to another place like earth" may be difficult to read. The structure of this sentence may need to be clarified.			S4. Is Sentence too difficult S9. Weak sentence start: ...
Logic	142	Tone/ Complexity/ General				S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Logic	156	Tone/ Vagueness/ Hedger	"A little" expresses uncertainty and should be used only when this stance is appropriate.			
Logic	208	Grammar/ Major/ Comma				P2. Is comma needed after ...
Logic	209	Clarity/ Theme	Sentences with too many subordinate ideas can be difficult to read. Can you clarify?			S4. Is Sentence too difficult
Logic	216	Clarity/ Readability/ Transitions	"But instead" may contradict itself or contain unnecessary 'transitional' words like "however" and "yet."			
Logic	221	Clarity/ Ambiguity Clarity/ Clarity/ Usage Related Clarity/ Readability/ Difficulty Clarity/ Readability/ Flow Clarity/ Readability/ Rhythm	The words around "is less farms" may be difficult to read. Are they used correctly?			G9. Is ... being used correctly G11. Is ... correct S4. Is Sentence too difficult S15. Is this ambiguous: ...
Logic	222	Grammar/ Missing Words	The words "be effected" may be used incorrectly, or the following words may be unclear.			G9. Is ... being used correctly G11. Is ... correct
Logic	232	Tone/ Vagueness/ Weak Conditional	"Can" weakens the conditional "if."			
Logic	234	Clarity/ Ambiguity	Should this sentence be read as "hitler also" or "also came." There may be several ways of interpreting this wording.			S15. Is this ambiguous: ...
Logic	262	Clarity/ Readability/ Difficulty				S4. Is Sentence too difficult

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Logic	266	Clarity/ Readability/ Difficulty	The words around "huminty" may be unclear, part of a special phrase, or what is left of an ellipsis of a phrase or clause. See 'Tutorial' for more information.			S4. Is Sentence too difficult
Logic	270	Clarity/ Clarity/ Meaning Related	Verb phrases like "should not be conducted only" may be difficult to understand. Could this one be simplified?			S4. Is Sentence too difficult
Logic	285	Clarity/ Wordiness/ Run-on/ Fused	It may be difficult to read from "the student will not take any more attention to they" to "so do it is so difficult." Is this a fused or run-on sentence? Is a subordinator such as 'that' missing? Is your point clear?	25. This appears to be a run-on sentence.		G3. Split into 2 sentences S4. Is Sentence too difficult
Logic	289	Clarity/ Readability/ Interruptions	The words between "methods" and "are" interrupt the flow between the subject and the verb. This sentence may read better if some or all of these words are moved elsewhere.			
Logic	400	Clarity/ Clarity/ Usage Related	The use of "nature and showy manner" and "was" may be unclear or overly complex. "nature and showy manner" and "was" may be part of an unclear subject-verb relationship. Could you clarify the topic of this sentence?			S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Logic	401	Clarity/ Readability/ Flow	Around "however some have prejudices against the exploration and see only the disadvantages of it" the sentence loses its flow. Can you clarify?			
Logic	402	Clarity/ Readability/ Flow	Around "must perform" the sentence loses its flow. Can you clarify?			
Logic	404	Clarity/ Readability/ Flow	This sentence does not flow well. Can you clarify?			
Logic	405	Clarity/ Readability/ Flow	This sentence does not flow well. Can you clarify?			

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Logic	406	Clarity/ Readability/ Flow	This sentence may be difficult to understand. "in which" and the preceding comma are part of the confusion. Can you clarify?			S4. Is Sentence too difficult
Logic	407	Clarity/ Readability/ Flow	This sentence may be difficult to understand. The punctuation around "each individual has their position or office" may be part of the confusion. Is this a fused or run-on sentence? Can you clarify?	25. This appears to be a run-on sentence.		G3. Split into 2 sentences S4. Is Sentence too difficult
Logic	408	Clarity/ Wordiness/ Introductions	The introductory part of this sentence may be unclear or too long for this sentence. Can you clarify, shorten or punctuate better?			S9. Weak sentence start: ...
Logic	409	Clarity/ Clarity/ Usage Related	The words following "arose" may be unclear. Has something been added or left out? Can you clarify?			
Logic	410	Style/ Word Selection/ General	The use of "cant" and "understand" may be unclear. Are they related properly? Can you clarify or use different words?		8. Homonyms	G12. Wrong word. Replace ... by ...
Logic	412	Clarity/ Readability/ Difficulty	Your point may not be clear as your reader proceeds from "if teachers are able to arose their interest by making the learning process fun and enjoyable" to "perharps students attitude might changed." Is this a fused or run-on sentence? Could you clarify?	25. This appears to be a run-on sentence.		G3. Split into 2 sentences S4. Is Sentence too difficult
Logic	413	Clarity/ Clarity/ Usage Related	The use of "affairs" in this sentence may be unclear. Is there a word missing in front of it?			
Precision	068	Clarity/ Clarity/ Vague Referents	Is it clear to what or whom "this" refers? Do you want to be more definite? Is its meaning clear?			
Precision	131	Tone/ Vagueness/ General	"Everything" may be vague. Could you use a more forceful word?	63. Unnecessary modifier. Omit or use more precise expression.		

BEST COPY AVAILABLE

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Precision	133	Tone/ Vagueness/ Weak	Weak words like 'big' do not convey much useful information in this context. Should a more descriptive word be used?	28. Weak modifier. Consider using a more precise expression. 70. Weak or unnecessary modifier consider using ... alone		S17. Weak: ... S18. Weak: Replace ... by ... U6. Consider using: ... U19. Is the modifier correct for absolute word? ...
Precision	171	Tone/ Vagueness/ General	Could you be more specific than "everything?"	23. Vague quantifier. Be more specific or try		
Precision	180	Tone/ Vagueness/ Unclear	The topic "factor" is weak. Can you use another word that is more descriptive?			
Precision	188	Clarity/ Readability/ Difficulty	The phrase "will only feel motivated or anticipated" has a lot of words or may be hard to read. Is there a simpler way to make your point?			S4. Is Sentence too difficult S12. Can simpler terms be used S13. Replace ... by simpler ... S13. Replace ... form of simpler ...
Precision	203	Grammar/ Usage/ Incorrect		72. Word usage consider ... instead.		G9. Is ... being used correctly G11. Is ... correct
Precision	214	Clarity/ Theme	The topic "several" and focus "things" are both vague. Should you be more specific with the main section of this sentence?			
Precision	227	Tone/ Vagueness/ General	"One" may not be the best subject, especially when used with "is" as a verb.			
Precision	231	Clarity/ Insufficient Information				
Precision	233	Tone/ Vagueness/ General	"Example" conveys little information. Could a more informative or specific word be found?	23. Vague quantifier. Be more specific or try		
Precision	247	Clarity/ Sprawl	There are a lot of prepositional phrases in this sentence. It may be unclear or difficult to read.	4. Consider revising. Long sequences of prepositional phrases can be confusing.		S4. Is Sentence too difficult
Precision	248	Clarity/ Clarity/ Vague Referents	The use of words such as "they, each, them, he ..." may cause this sentence to be vague. Could you be more specific?			

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Precision	249	Clarity/ Nominalizations	The actions in this sentence could be more directly expressed. "Nominalized" words such as "chosen" and "decision" express in nouns the actions that are normally expressed by verbs and adjectives. See 'Tutorial' for details.			
Precision					2. Vague adverb	
Punctuation	016	Grammar/ Major/ Comma	There may be a structural problem in this sentence. The words around "can get" may be the source of the problem. Is a comma needed at some point?	47. Consider adding a comma after ... . 79. Avoid using two superlatives not separated by a comma.		P2. Is comma needed after ...
Punctuation	020	Style/ Writing/ Excessive Punctuation	This sentence is heavily punctuated. Are all these punctuation marks necessary?			
Punctuation	038	Grammar/ Punctuation/ General		44. Consider deleting the period after ... 54. deleting the period after 67. Consider deleting this punctuation mark		
Punctuation	058	Clarity/ Readability/ Difficulty				S4. Is Sentence too difficult
Punctuation	060	Grammar/ Major/ Commas				P2. Is comma needed after ...
Punctuation	063	Grammar/ Punctuation/ Capitalization			42. Capitalization	C1. Unusual capitalization: ... C2. Do not capitalize: ... C3. Capitalize: ... C4. Should ... be capitalized
Punctuation	099	Grammar/ Sentence Structure/ Interrogative				P1. Is question mark missing
Punctuation	100	Grammar/ Sentence Structure/ Declaratives	Should this sentence end with a period?		49. End of sentence punctuation	
Punctuation	102	Grammar/ Major/ Comma	Introductory words like "in addition" are often followed by a comma.			P2. Is comma needed after ...

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Punctuation	105	Clarity/ Wordiness/ Run-on/ Fused	This sentence may have more than one main idea. You may need a semicolon to separate them, or you may need to simplify the sentence. Check the wording around "my friends and I are very competious" and "we are rivals."	25. This appears to be a run-on sentence.		G3. Split into 2 sentences
Punctuation	218	Grammar/ Major/ Semicolon	The semicolon after "and so on" may be inappropriate in this context. The following words do not seem to have a main idea.	32. The semicolon seems inappropriate in this context.		P4. Semucolons separate independent clauses
Punctuation	225	Grammar/ Major/ Comma	The comma after "we" may need to be removed, or the surrounding words clarified.			
Punctuation	238	Grammar/ Major/ Comma	"And" seems to come between two main ideas. If so, you may want a comma before "And."			
Punctuation				5. The abbreviation ... is not set off by the correct punctuation. 41. The abbreviation ... should be preceded by a comma.		
Punctuation				6. Consider changing or deleting the double quotation mark.	14. Quotation marks 52. Quotation misuse	S20. Single word enclosed by quotes P8. was this quote opened
Punctuation				11. The quoted material appears to be improperly punctuated.		
Punctuation				16. This sentence appears to need a double quotation mark.		P7. Is this quote closed
Punctuation				21. Put the period inside the quotation marks unless they set off special terms.		
Punctuation				27. Consider putting the comma inside the quotation mark.		
Punctuation				33. Consider adding a space after this punctuation mark.		

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Punctuation				48. Consider deleting the space before this punctuation mark. 53. spacing around this punctuation mark 78. Consider deleting the space after this punctuation mark.		P14. Remove space before punctuation
Punctuation				38. This punctuation combination is unusual.	12. Punctuation Usage	P13. Is this punctuation correct
Punctuation				51. need a right parenthesis 57. Consider putting this punctuation mark outside the parenthesis. 80. Consider putting this punctuation mark inside the parenthesis	15. Unbalanced parentheses	P9. Is this bracket closed P10. Was this bracket opened P11. Is this parenthesis closed P12. Was this parenthesis opened
Punctuation						P6. Reversed punctuation
Punctuation				37. Avoid using dashes too frequently in a single sentence.		
Relation	009	Clarity/ Ambiguity				S15. Is this ambiguous ...
Relation	010	Grammar/ Usage/ Determiners	"Its a" may have too many words such as "the," "a," "some", "any," "these," "that"... Could one be removed, or could this section be restated? Is there a comma missing between them?			P3. Is comma missing after ...
Relation	021	Style/ Optional Usage/ Commas	A comma may be needed between "cultural" and "very" to clarify your meaning. See "Tutorials" for a detailed explanation.			
Relation	023	Clarity/ Readability/ Difficulty				S4. Is Sentence too difficult
Relation	024	Tone/ General/ Similar Modifiers	"Gradually" and "gradually" sometimes cause confusion when used together. Should one be removed? Should they be coordinated?		9. Commonly confused 46. Similar words	U19. Is the modifier correct for absolute word? ...
Relation	054	Clarity/ Insufficient Information	This sentence may have a word missing after "easier," a faulty coordination of phrases, or an unclear ellipsis. Can you clarify?		11. Ellipsis Mark 48. Ellipse usage	

BEST COPY AVAILABLE

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Relation	072	Style/ Word Selection/ Best Wording	Is "mine of" the best wording? If so, is "of" where it belongs?			
Relation	080	Clarity/ Ambiguity				S15. Is this ambiguous: ...
Relation	082	Clarity/ Complexity/ General Relationships	"Will" and "depends" seem to be verb forms used incorrectly. Is a word missing between them?			G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S5. Use verb form. Replace ... by ...
Relation	090	Grammar/ Ambiguity				S15. Is this ambiguous: ...
Relation	107	Grammar/ Sentence Structure/ Position	"Anything" usually follows a word like "not." See 'Tutorial' for more information.			
Relation	113	Clarity/ Clarity/ Vague Referents	It may not be clear to whom or what "his or her" refers.			
Relation	130	Clarity/ Clarity/ Vague Referents	Is it clear to what "another" refers? Do you want to be more specific?			
Relation	170	Tone/ Vagueness/ Unclear		23. Vague quantifier. Be more specific or try ...		
Relation	187	Clarity/ Sprawl	The amount of detail in "for the teacher to near behind the student" may obscure your main point. Could part of it be moved to another place in the sentence? Could some of the detail be deleted?			
Relation	189	Clarity/ Theme	This sentence has a lot of descriptive information in it. It may not be clear what to focus on.			
Relation	204	Grammar/ Coordination	The coordination in "how much or how little" should be avoided.			
Relation	223	Grammar/ Major/ Comma	"Because ... has a strong link with the environment and exposure to nature" may be used incorrectly. There may need to be a comma before and after it, or the surrounding words clarified.			G9. Is ... being used correctly G11. Is ... correct.

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Relation	225	Grammar/ Major/ Comma	The comma after "from the fact" is not required in this context, unless its removal would make the sentence ambiguous.			
Relation	243	Clarity/ Sprawl	There are a lot of modifying elements in this sentence. It may not be clear what they are modifying, or there may be too much additive information.			
Relation	265	Grammar/ Usage/ General Relationship	Does "our" belong with "our live"? If so, "our live" and the following words may be unclear.			
Relation	414	Clarity/ Clarity/ Time Related	It may be difficult to place the time of the actions in this sentence. Words such as "since" and "are" are used in complicated ways. Can you clarify?			S4. Is Sentence too difficult
Relation				24. Consider using ... as the restrictive relative pronoun.		
Surface	105	Clarity/ Wordiness/ Run-on/ Fused	This sentence may have more than one main idea. If you are indirectly quoting someone, this may be correct. Otherwise, you may need a semicolon to separate them. Check the wording around "in the big picture, it is true" and "that outrageous behavior will reflect the standards of society as a whole."	25. This appears to be a run-on sentence.		G3. Split into 2 sentences
Surface	123	Grammar/ Spelling Spell Check		14. The word ... may be misspelled. 64. Consider ... instead	27. Single-word spelling 28. Split-word spelling 29. Similar spelling 44. Spelling	U14. Is this a word? ... U15. No a word: ...
Surface	127	Grammar/ Usage/ Non-Standard English	"Layed" is not standard English.	74. This word may not be used with this contraction		
Surface	236	Clarity/ Readability/ Difficulty	This sentence may take several readings to be understood. Should it be rewritten?			S4. Is Sentence too difficult

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Surface	267	Grammar/ Spelling/ Automatic Connections	The misspelled word "aparently" has been corrected to "apparently." If you agree with this correction, then there is nothing more to do.	17. Open Vs closed spelling. Consider ... instead. 22. The preferred spelling of ... is		U16. Not a word. Replace ... by ...
Transition	036	Clarity/ Readability/ Position	This sentence might be easier to read if "in which we as puertorriquenos live in there is a very small chance of that action" were in the first part of the sentence.			
Transition	059	Clarity/ Wordiness/ Introductions	Is "despite man's ability to be independent" the introductory part of this sentence? If so, the introduction may be too long for this sentence. You may want to re-organize this sentence.			S9. Weak sentence start. ...
Transition	185	Style/ Word Position/ General	"At the same time" may read better if moved to the front of the clause. See 'Tutorial' for more information.			
Unity	075	Clarity/ Ambiguity				S15. Is this ambiguous ...
Unity	110	Grammar/ Usage/ Split Infinitives	(Split Infinitive) The words between "to" and "lie" do not belong there. They may go before "to" or after "lie" or may need to be removed.	75. The sequence ... may be a split infinitive.	40. Infinitive usage 51. Split infinitive	S2. Split infinitive ...
Unity	182	Clarity/ Wordiness/ Excessive Info				
Unity	186	Clarity/ Sprawl				
Unity	190	Clarity/ Clarity/ Vague Referents	"They" can refer to more than one noun here. Make sure that it is clear which noun it refers to.	23. Vague quantifier. Be more specific or try		
Unity	237	Clarity/ Read/ Flow	The words "because when" coming one after the other may be difficult to understand.			S4. Is Sentence too difficult
Unity	239	Clarity/ Clarity/ Usage Related				
Unity	251	Clarity/ Clarity/ Vague Referents	Does "with such idealogy then" refer to "enter"? It may not be clear and could be interpreted in more than one way.	23. Vague quantifier. Be more specific or try		

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Unity	252	Clarity/ Clarity/ Misplaced Modifiers	The relationship of the introductory phrase "because by so many people's effort" to the following words may be unclear.	49. Rephrase to replace this dangling modifier with a more specific phrase.		
Unity	253	Grammar/ Coordination	If "days" and "child" are in a series, they should be of the same type. Are they? If they are not in a series, the wording between them may be too complex.			
Usage	005	Grammar/ Usage/ Determiners				G6. Replace A by AN G7. Replace AN by A
Usage	022	Grammar/ Plurals & Possessives/ Possessive Needed	The possessive form of "boys" may be needed here, unless "boys" is a modifier or part of a special phrase.		4. Possessive Form 39. Possessive Usage	G10. Should ... be possessive
Usage	026	Grammar/ Usage/ Incorrect				G9. Is ... being used correctly G11. Is ... correct
Usage	028	Clarity/ Usage Related	"Willingly" and "go" don't seem to belong together.			
Usage	043	Grammar/ Verbs/ Usage	"Are not" cannot normally be used with another word ("be") of the same type. Has a word been deleted?			G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S5. Use verb form. Replace ... by ...
Usage	053	Grammar/ Modification/ Incorrect	"One" does not seem to match "sets." Do they belong together? Are they part of a special phrase? Has a word such as "that" been deleted? Is there a missing comma?			G9. Is ... being used correctly G11. Is ... correct
Usage	064	Grammar/ Misplaced Words	"There" may be used incorrectly here. Should an adjective form be used, or is there a word missing?			G9. Is ... being used correctly G11. Is ... correct
Usage	069	Grammar/ Usage/ Incorrect			43. Usage in question	G9. Is ... being used correctly G11. Is ... correct
Usage	076	Grammar/ Sentence/ Structure/ Position	Is "the ability" in the most effective position? If so, is it properly connected to another part of the sentence? Is it clear, or does it contain too much additional information?			

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Usage	077	Style/ Word Position/ Prepositions	This sentence ends with the preposition "before." Some audiences may find this too informal. See 'Tutorial' for some better alternatives.			S10. Sentence ends with preposition
Usage	078	Clarity/ Wordiness/ Run-on/ Fused		25. This appears to be a run-on sentence.		G3. Split into 2 sentences
Usage	083	Clarity/ Insufficient Information	"Set" often takes one or more modifiers not found here. See 'Tutorial' for additional information.			
Usage	084	Grammar/ Usage/ Incorrect	The personal pronoun "us" may be the wrong form of pronoun in this context. See 'Tutorial' for some better alternatives.	45. The pronoun ... should come last in a series of conjoined nouns.	6. Pronoun Usage	G5. Wrong pronoun, replace ... by ...
Usage	085	Style/ Word Selection/ Double Negatives	"From not" contains more than one word with a negative force. Can this be stated in a positive way?	9. Avoid using double negatives.	32. Double negative	
Usage	088	Grammar/ Usage/ Incorrect				G9. Is ... being used correctly G11. Is ... correct.
Usage	089	Grammar/ Verbs/ Usage				G4. Wrong verb, replace ... by ... G8. Is ... the correct form of the verb S5. Use verb form. Replace ... by ...
Usage	096	Grammar/ Ambiguity				S15. Is this ambiguous ...
Usage	097		"Which" is best used to introduce additional information. Is this the case here? See 'Tutorial' for some better alternatives.			
Usage	121	Tone/ General/ Archaic		1. Archaic expression. Consider ... instead.	21. Archaic	U3. Archaic ... U4. Archaic. Replace ... by ...
Usage	149	Tone/ General/ Usage	"Assured" is often misused.		37. Often misused or confused.	
Usage	151	Tone/ General/ Overused	"It goes without saying that" tends to be overused and may not be necessary in this sentence.	19. Overused. Use sparingly.		S19. Overused: ...
Usage	164	Tone/ General/ Usage				
Usage	226	Grammar/ Usage/ Incorrect	If "that" refers to "nurse," it might need to be replaced by "who/whom." If not, the referent for "that" may be unclear.			G9. Is ... being used correctly G11. Is ... correct. G12. Wrong word. Replace ... by ...

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Usage	235		Should the "er" or "est" form of "friendly" be used instead of "more friendly?"	58. Consider rephrasing with ... 66. Use "different form" or rephrase using a more specific comparative	50. Comparative usage	
Usage	261	Clarity/ Clarity/ Usage Related				
Usage	271	Grammar/ Usage/ General Relation	"May" does not seem appropriate following "are." Should it be moved to another position or replaced with another word?			
Usage	272	Grammar/ Usage/ General Relation				
Usage	273	Grammar/ Usage Incorrect				G9. Is ... being used correctly G11. Is ... correct
Usage	275	Grammar/ Usage/ General Relation				
Usage	279	Clarity/ Insufficient Information	The use of "unattainable" may not be clear. Would it be better to replace "unattainable" with another noun, add a noun after it, or move "unattainable" in front of the noun that it modifies?	62. Unless ... modifies the preceeding noun, try ...		G12. Wrong word. Replace by ...
Usage	283	Grammar/ Usage/ Incorrect				G9. Is ... being used correctly G11. Is ... correct
Usage	284	Clarity/ Clarity/ Usage Related	"That do to solving the problems of society" may be incorrect or unclear when following "can." Could you clarify? Should "That do to solving the problems of society" be moved to another sentence? Is "That do to solving the problems of society" the correct wording? Is a comma needed after "can?"			G9. Is ... being used correctly G11. Is ... correct
Usage				3. Consider rephrasing with a form of ...		
Usage				12. Consider ... instead of ... . 60. Consider ... instead of ... 61. Considre ... instead of ...		U6. Consider using: ...
Usage			73. Preposition consider "outside" unless you mean 'excepting'	31. Unless this means ..., use ... 55. unless you are stressing the alternatives		

BEST COPY AVAILABLE

Category	Error Number in Power Edit	Error Description in PowerEdit	Error Message in PowerEdit	Error Message in CorrectGrammar	Error Message in Grammatik	Error Message in RightWriter
Usage				34. Preposition usage. Delete ... or rephrase with a form of ...	5. Preposition	
Usage					1. Adverb	
Usage					10. Doubled word or punctuation	
Usage					25. Questionable Usage	
Usage						U9. Is this justified: ...
Usage						U10. Is this explained: ...
Usage						U18. Consider rephrasing
Usage						U22. User Flagged Word: ...

75

Appendix C
Essay Analysis Data Record Format

Each essay analysis produced a record containing the following data.

essay identifier
first reader grade
second reader grade
word count for the essay
sentence count for essay
number of words that PowerEdit could not analyzer for and essay

total number of balance errors found by PowerEdit
total number of balance errors found by Grammatik
total number of balance errors found by CorrectGrammar
total number of balance errors found by RightWriter

total number of cohesion errors found by PowerEdit
total number of cohesion errors found by Grammatik
total number of cohesion errors found by CorrectGrammar
total number of cohesion errors found by RightWriter

total number of concision errors found by PowerEdit
total number of concision errors found by Grammatik
total number of concision errors found by CorrectGrammar
total number of concision errors found by RightWriter

total number of discourse errors found by PowerEdit
total number of discourse errors found by Grammatik
total number of discourse errors found by CorrectGrammar
total number of discourse errors found by RightWriter

total number of elegance errors found by PowerEdit
total number of elegance errors found by Grammatik
total number of elegance errors found by CorrectGrammar
total number of elegance errors found by RightWriter

total number of emphasis errors found by PowerEdit
total number of emphasis errors found by Grammatik
total number of emphasis errors found by CorrectGrammar
total number of emphasis errors found by RightWriter

total number of grammar errors found by PowerEdit
total number of grammar errors found by Grammatik
total number of grammar errors found by CorrectGrammar
total number of grammar errors found by RightWriter

total number of logic errors found by PowerEdit
total number of logic errors found by Grammatik
total number of logic errors found by CorrectGrammar
total number of logic errors found by RightWriter

total number of precision errors found by PowerEdit
total number of precision errors found by Grammatik
total number of precision errors found by CorrectGrammar

total number of precision errors found by RightWriter

total number of punctuation errors found by PowerEdit
total number of punctuation errors found by Grammatik
total number of punctuation errors found by CorrectGrammar
total number of punctuation errors found by RightWriter

total number of relation errors found by PowerEdit
total number of relation errors found by Grammatik
total number of relation errors found by CorrectGrammar
total number of relation errors found by RightWriter

total number of surface errors found by PowerEdit
total number of surface errors found by Grammatik
total number of surface relation errors found by CorrectGrammar
total number of surface errors found by RightWriter

total number of transition errors found by PowerEdit
total number of transition errors found by Grammatik
total number of transition relation errors found by CorrectGrammar
total number of transition errors found by RightWriter

total number of unity errors found by PowerEdit
total number of unity errors found by Grammatik
total number of unity relation errors found by CorrectGrammar
total number of unity errors found by RightWriter

total number of usage errors found by PowerEdit
total number of usage errors found by Grammatik
total number of usage relation errors found by CorrectGrammar
total number of usage errors found by RightWriter